

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 23:02:20 ; Search time 2628 Seconds

(without alignments)
4152.798 Million cell updates/sec

Title: US-09-936-737A-1

Sequence: 375
1 atgaagattcttctgattc.....atgaagtgtgatgaattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :
1: gb_ba :
2: gb_hg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_vi :
30: em_hg_hum :
31: em_hg_inv :
32: em_hg_other :
33: em_hg_mus :
34: em_hg_pln :
35: em_hg_rod :
36: em_hg_mam :
37: em_hg_vrt :
38: em_sy :
39: em_hg_hum :
40: em_hggo_mus :
41: em_hggo_other :

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|------------|--------------------|
| 1 | 375 | 100.0 | 375 | 6 | AX036541 | AX036541 Sequence |
| 2 | 48.6 | 13.0 | 189967 | 2 | AL833802 | AL833802 Mus muscu |
| 3 | 48.6 | 13.0 | 206914 | 10 | AL807815 | AL807815 Mouse DNA |
| 4 | 46.4 | 12.4 | 40090 | 3 | AC006611 | AC006611 Caenorhab |
| 5 | 46.4 | 12.4 | 299670 | 2 | AC006845 | AC006845 Caenorhab |
| 6 | 45.8 | 12.2 | 4458 | 2 | AC115614 | AC115614 Dictyoste |
| 7 | 45.8 | 12.2 | 49999 | 6 | AX015902 | AX015902 Sequence |
| 8 | 45.8 | 12.2 | 49999 | 6 | AX015908 | AX015908 Sequence |
| 9 | 45.8 | 12.2 | 119191 | 9 | AP004782 | AP004782 Homo sapi |
| 10 | 45.8 | 12.2 | 228283 | 10 | AF312994 | AF312994 Mus muscu |
| 11 | 45.2 | 12.1 | 20796 | 3 | AF216973 | AF216973 Drosophil |
| 12 | 45.2 | 12.1 | 87767 | 2 | AC014497 | AC014497 Drosophil |
| 13 | 45.2 | 12.1 | 132000 | 2 | AC016976 | AC016976 Dictyoste |
| 14 | 45.2 | 12.1 | 167062 | 2 | AC007624 | AC007624 Drosophil |
| 15 | 45.2 | 12.1 | 181771 | 3 | AC008340 | AC008340 Drosophil |
| 16 | 44.8 | 11.9 | 26671 | 3 | CER11A5 | 283122 Caenorhabd |
| 17 | 44.4 | 11.8 | 3230 | 9 | BC002875 | BC002875 Homo sapi |
| 18 | 44.4 | 11.8 | 86945 | 2 | BC010457 | BC010457 Homo sapi |
| 19 | 44.4 | 11.8 | 86945 | 2 | AC002490 | AC002490 Homo sapi |
| 20 | 44.4 | 11.8 | 153477 | 2 | AC006278 | AC006278 Plasmodiu |
| 21 | 44.4 | 11.8 | 154470 | 2 | AC027820 | AC027820 Homo sapi |
| 22 | 44.4 | 11.8 | 195770 | 2 | AC091153 | AC091153 Homo sapi |
| 23 | 44.2 | 11.6 | 247175 | 2 | AC129593 | AC129593 Mus muscu |
| 24 | 43.6 | 11.6 | 152409 | 2 | PFMAL1P2_1 | Continuation (2 of |
| 25 | 43.6 | 11.6 | 183767 | 2 | PFMAL1P1 | AL031744 Plasmodiu |
| 26 | 43.6 | 11.6 | 234295 | 2 | PFMAL1P2 | AL807765 Mus muscu |
| 27 | 43.6 | 11.6 | 98734 | 2 | AL714007 | AL714007 Mus muscu |
| 28 | 43.4 | 11.6 | 143716 | 2 | AL627248 | AL627248 Dario rer |
| 29 | 43.4 | 11.6 | 168083 | 9 | AC092376 | AC092376 Homo sapi |
| 30 | 43.2 | 11.5 | 977 | 11 | CNS061DX | AL400171 T7 end of |
| 31 | 43.2 | 11.5 | 139665 | 2 | AP000865 | AP000865 Homo sapi |
| 32 | 43.2 | 11.5 | 110000 | 2 | PFMAL1P1_1 | Continuation (2 of |
| 33 | 43 | 11.5 | 162261 | 2 | AC119001 | AC119001 Rattus no |
| 34 | 42.8 | 11.4 | 142439 | 9 | AC008749 | AC008749 Homo sapi |
| 35 | 42.8 | 11.4 | 59762 | 8 | AB023032 | AB023032 Arabidops |
| 36 | 42.6 | 11.4 | 173016 | 2 | AC019123 | AC019123 Homo sapi |
| 37 | 42.6 | 11.4 | 262428 | 2 | AC096326 | AC096326 Rattus no |
| 38 | 42.6 | 11.3 | 2783 | 3 | AF093702 | AF093702 Plasmodiu |
| 39 | 42.4 | 11.3 | 49144 | 2 | AC131380 | AC131380 Strongylo |
| 40 | 42.4 | 11.3 | 115758 | 9 | AC104634 | AC104634 Homo sapi |
| 41 | 42.4 | 11.3 | 198516 | 2 | AC102659 | AC102659 Mus muscu |
| 42 | 42.4 | 11.3 | 202521 | 2 | AL773546 | AL773546 Mus muscu |
| 43 | 42.4 | 11.3 | 234112 | 3 | PFMAL1P2 | AL035475 Plasmodiu |
| 44 | 42.4 | 11.3 | 388 | 8 | PAB296740 | AJ296740 Picea abi |
| 45 | 42.2 | 11.3 | | | | |

ALIGNMENTS

RESULT 1
AX036541
LOCUS
DEFINITION
Sequence 1 from Patent WO0056885.
AX036541
VERSION
AX036541.1 GI:11226151
KEYWORDS
SOURCE
ORGANISM
medicinal leech.
Hirudo medicinalis
Enkaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
Arynchobdellida; Hirudiniiformes; Hirudiniidae; Hirudo.
REFERENCE
1 (bases 1 to 375)
Hemberger, U., Scheuble, B., Strittmatter, W., Hofmann, U., Fotev, Z.
and Guesow, D.
Protein for blocking platelet adhesion
TITLE

Pred. No. is the number of results predicted by chance to have a

adp. date before 1.0.1
Pub date after 1.0.1

JOURNAL

Patent: WO 0056885-A 1 28-SEP-2000;
HEMBERGER JUERGEN (DE) ; MERCK PATENT GMBH (DE) ; SCHEIBLE BERNHARD
(DE) ; STRITTMATTER WOLFGANG (DE) ; HORMANN UWE (DE) ; FOTEV ZISI
(DE) ; GIBSSON DETLEF (DE)
Location/Qualifiers

FEATURES

Source 1..375

CDS

/organism="Hirudo medicinalis"
/db_xref="taxon:6421"
<64..375
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16481.1"
/db_xref="GI:11226152"
/translation="EEREEDCWTFYANKRYNDPDKSPKSSDLECKKTEKTEYCTIV
FEDTVNKECYNNVDDGELDQEFYVDENTENTYLTDCBKGDAAGAGDESDVEDE
D"

BASE COUNT 123 a 62 c 85 g 105 t
ORIGIN

Query Match 100.0%; Score 375; DB 6; Length 375;
Best Local Similarity 100.0%; Pred. No. 6, 6e-86;

Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTATTTCTTGATTTCTTCCCTTGGCTCGCAGCTTGATCTCACTACTTCT 60
1 ATGAAGTATTTCTTGATTTCTTCCCTTGGCTCGCAGCTTGATCTCACTACTTCT 60
QY 61 TCAGAACAGCTGAGAGATTTGTGACGCTTTAGCCGACGAGAAATATACAGCTTCGAT 120
61 TCAGAACAGCTGAGAGATTTGTGACGCTTTAGCCGACGAGAAATATACAGCTTCGAT 120
QY 121 AAATCTTTTAAAGATGCTCTGATCTTGACGAGTCAAAAACATGTTTCAAGACGGAG 180
121 AAATCTTTTAAAGATGCTCTGATCTTGACGAGTCAAAAACATGTTTCAAGACGGAG 180
Db 121 AAATCTTTTAAAGATGCTCTGATCTTGACGAGTCAAAAACATGTTTCAAGACGGAG 180
121 TACTGCTACATCGTTTGTGAAGACAGCGCTCAACAGAGTCTTACATGCTGTGAT 240
181 TACTGCTACATCGTTTGTGAAGACAGCGCTCAACAGAGTCTTACATGCTGTGAT 240
Db 181 TACTGCTACATCGTTTGTGAAGACAGCGCTCAACAGAGTCTTACATGCTGTGAT 240
241 GGTGAAGAGTTAGACCAAGAAAATTTGTGTGACGAGAAACTTCAGGAAATTTTGG 300
241 GGTGAAGAGTTAGACCAAGAAAATTTGTGTGACGAGAAACTTCAGGAAATTTTGG 300
QY 301 ACAGACTGCGAGGTTAAAGATGCAAGTATGCGGACGAGTACAGTACAGTACAGTAA 360
301 ACAGACTGCGAGGTTAAAGATGCAAGTATGCGGACGAGTACAGTACAGTACAGTAA 360
Db 301 ACAGACTGCGAGGTTAAAGATGCAAGTATGCGGACGAGTACAGTACAGTACAGTAA 360
QY 361 GTTGATGAAGATTAA 375
361 GTTGATGAAGATTAA 375
Db 361 GTTGATGAAGATTAA 375

RESULT 2

AL833802

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL833802 189967 bp DNA linear HTG 17-JUL-2002
Mus musculus chromosome 4 clone RP23-17A4, *** SEQUENCING IN
PROGRESS *** 13 unordered pieces.
AL833802.3 GI:21911787
HTG: HTGS_PHASE1.
SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 189967)
Plumb, B.
Direct Submission
Submitted (16-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jul 19, 2002 this sequence version replaced gi:21738804.
----- Genome Center
Center: Wellcome Trust Sanger Institute

FEATURES

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM17A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 186623 bases at least Q40
Consensus quality: 187480 bases at least Q30
Consensus quality: 188086 bases at least Q20
Insert size: 188767; sum-of-contrigs
Insert size: 215919; 10.9% error; agarose-fp
Quality coverage: 5.28x in Q20 bases; sum-of-contrigs Quality
coverage: 4.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 8454: contig of 8454 bp in length
* 8455 8554: gap of 100 bp
* 8555 19307: contig of 10753 bp in length
* 19308 19407: gap of 100 bp
* 19408 27097: contig of 7690 bp in length
* 27098 27197: gap of 100 bp
* 27198 40822: contig of 13625 bp in length
* 40823 40922: gap of 100 bp
* 40923 48761: contig of 7839 bp in length
* 48762 48861: gap of 100 bp
* 48862 94932: contig of 46071 bp in length
* 94933 95032: gap of 100 bp
* 95033 109692: contig of 14660 bp in length
* 109693 109792: gap of 100 bp
* 109793 122076: contig of 12284 bp in length
* 122077 122176: gap of 100 bp
* 122177 130769: contig of 8593 bp in length
* 130770 130869: gap of 100 bp
* 130870 146267: contig of 15398 bp in length
* 146268 146367: gap of 100 bp
* 146368 151984: contig of 5617 bp in length
* 151985 152084: gap of 100 bp
* 152085 184112: contig of 32028 bp in length
* 184113 184212: gap of 100 bp
* 184213 189967: contig of 5755 bp in length.
Location/Qualifiers
1..189967
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-17A4"
/clone_1bp="RP23-17A4"
1..8454
/note="assembly-fragment:01243
fragment_chain:1
clone_end:SP6
vector_side:left"
8555..19307
/note="assembly-fragment:00590
fragment_chain:1"
19408..27097
/note="assembly-fragment:00259
fragment_chain:1"
27198..40822
/note="assembly-fragment:01468
fragment_chain:1"
40923..48761
/note="assembly-fragment:00793
fragment_chain:2"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

misc_feature      48862..94932
                  /note="assembly-fragment:00456
                  fragment_chain:2"
misc_feature      95033..109692
                  /note="assembly-fragment:00530
                  fragment_chain:2"
misc_feature      109793..122076
                  /note="assembly-fragment:01290
                  fragment_chain:2"
misc_feature      122177..130769
                  /note="assembly-fragment:01504.0"
misc_feature      130870..146267
                  /note="assembly-fragment:01504.1"
misc_feature      146368..151984
                  /note="assembly-fragment:00822
                  fragment_chain:3"
misc_feature      152085..184112
                  /note="assembly-fragment:00663
                  fragment_chain:3"
misc_feature      184213..189967
                  /note="assembly-fragment:00178
                  fragment_chain:3"
misc_feature      184213..189967
                  /note="assembly-fragment:00178
                  fragment_chain:3"
BASE COUNT      59392 a 34302 c 33745 g 61328 t 1200 others
ORIGIN
Query Match      13.0%; Score 48.6; DB 2; Length 189967;
Best Local Similarity 48.4%; Pred. No. 0.066;
Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 97 AACAGAAATATACAGACTTCGATTAATCTTTAAGAGTCTCTGATCTTGACGATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88182 AACAAAGCGCCGCAATATTGAAACAGTTCTTAACAATAAAGAACTTTGGGGGAATC 88241
QY 157 AAAAAACATGTTCAAGACGAGTCTGCTACATCTGTTTGAAGACACGCTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88242 ACCATTCATGACTTCAAGCTTATCTACAGCAATGATGATGACAGAAAGAAAGAG 88301
QY 217 GAATGTTACTACATGCTGTTGATGGTGAAGAGTACACCAAGAAATTTTGTGCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88302 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88361
QY 277 GAAACTTCAAGGAAATTTTTCAGACTCGAGGCTTAAGATGACAGTATATGCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88362 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88421
QY 337 GGTACAGGTGACGAGTCAAGTGAAGTGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88422 GATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88460

RESULT 3
AL807815/c      206914 bp      DNA      linear      ROD 22-AUG-2002
LOCUS          Mouse DNA sequence from clone RP23-207H16 on chromosome 4, complete
DEFINITION
ACCESSION      AL807815
VERSION        AL807815.3 GI:22474427
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS        Leongamornlert,D.
TITLE          Direct Submission
JOURNAL        Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail: enquiries@hinxton.sanger.ac.uk
COMMENT
Center: Genome Center
Center: Wellcome Trust Sanger Institute

```

Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems; such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-207H16 is from the RP23-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

```

FEATURES
source
Location/Qualifiers
1..206914
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-207H16"
/clone_11b="RP23-23"
BASE COUNT      64113 a 37887 c 37484 g 67430 t
ORIGIN
Query Match      13.0%; Score 48.6; DB 10; Length 206914;
Best Local Similarity 48.4%; Pred. No. 0.067;
Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 97 AACAGAAATATACAGACTTCGATTAATCTTTAAGAGTCTCTGATCTTGACGATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205163 AACAAAGCGCCGCAATATTGAAACAGTTCTTAACAATAAAGAACTTTGGGGGAATC 205104
QY 157 AAAAAACATGTTCAAGACGAGTCTGCTACATCTGTTTGAAGACACGCTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205103 ACCATTCATGACTTCAAGCTTATCTACAGCAATGATGATGACAGAAAGAAAGAG 205044
QY 217 GAATGTTACTACATGCTGTTGATGGTGAAGAGTACACCAAGAAATTTTGTGCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205043 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204984
QY 277 GAAACTTCAAGGAAATTTTTCAGACTCGAGGCTTAAGATGACAGTATATGCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 204983 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204924
QY 337 GGTACAGGTGACGAGTCAAGTGAAGTGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 204923 GATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204885

RESULT 4
AC006611      40090 bp      DNA      linear      INV 19-APR-2002
LOCUS          Caenorhabditis elegans cosmid C30F8, complete sequence.
DEFINITION
ACCESSION      AC006611
VERSION        AC006611.1 GI:4263327
KEYWORDS
SOURCE         Caenorhabditis elegans.
ORGANISM       Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS        I (bases 1 to 40090)

```

AUTHORS Waterston, R.
TITLE Genome sequence of the nematode *C. elegans*: a platform for
JOURNAL Investigating biology. The *C. elegans* Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
PUBMED 98696613
9851916

REFERENCE 2 (bases 1 to 40090)
AUTHORS Latreille, P., Twyman, B. and Wilson, R.
TITLE The sequence of *C. elegans* cosmid C30F8
JOURNAL Unpublished (2001)
AUTHORS 3 (bases 1 to 40090)
TITLE Waterston, R. H.
JOURNAL Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 40090)
AUTHORS Waterston, R. H.
TITLE Direct Submission
Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (14-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
6 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (18-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
7 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (12-FEB-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
8 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE 1 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (12-FEB-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
8 (bases 1 to 40090)
AUTHORS Waterston, R.
TITLE Direct Submission
Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: tw@nematoe.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
[www.wormbase.org/db/seq/sequence?name=C30F8;class=Sequence]

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46H11, 8300 bp overlap; the 3' cosmid is Y110A7A,

200 bp overlap. Actual start of this cosmid is at base position 197
of C30F8; actual end is at 40090 of C30F8.

NOTES:

Coding sequences below are the result of integration and manual
review of the following data: computer analysis using the program
GeneIndex (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yui Kohara
(http://www.ddb.jhu.edu/c-elegans/html/CE_INDEX.html) and the
elegans ORFome cloning project (http://world.dicf.harvard.edu/),
similarity to other proteins from Blast analyses
(http://blast.wustl.edu/), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:115-123, 2000), individual C. elegans Genbank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
Location/Qualifiers
1. 40090

FEATURES

source

/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="C30F8"
/complement(19565..21512)
/gene="C30F8.2"

gene

/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=C30F8.2;class=Sequence]"
complement(19565..21512)
20460..20617,20879..21057,21102..21273,21392..21512)

CDS

/note="coded for the following C. elegans cDNAs:
YK710g2.3, YK449g2.3, YK705g11.3, YK745d7.3, YK169b6.3,
YK809d10.3, YK807a04.3, YK702g3.3, YK105b9.3, YK105b9.5,
YK169b6.5, YK306e11.3, YK306e11.5, YK431f1.3, YK431f1.5,
YK449g2.5, YK479e6.5, YK859d05.5, YK710g2.5, YK702g3.5,
YK888b07.5, YK745d7.5, YK763e04.5, YK821a09.5, YK809d10.5,
YK773c05.5, YK705g11.5, YK754e11.5, YK807a04.5,
YK754e11.3, YK888b07.3, YK773c05.3, YK821a09.3"
/codon_start=1
/product="Hypothetical protein C30F8.2"
/protein_id="AAK85455.2"
/db_xref="GI:16259229"

misc_feature

misc_feature

gene

CDS

/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=C30F8.4a;class=Sequence]"
complement(21534..21535)
/note="SL1 trans-splice site; see YK809d10.5"
25607..32891
/gene="C30F8.4"
/note="similar to protein kinases (PFam: pkinase, score:
183.25); coded for by the following C. elegans cDNAs:
YK739d8.5, YK284c11.3, YK284c11.5, YK465a7.5, YK739d9.3,
YK649b5.5"
/codon_start=1
/product="Hypothetical protein C30F8.4a"
/protein_id="AAK85457.1"
/db_xref="GI:15150650"

| | | | | | |
|--------------------------|--|-------------|-----|--------|-----------------|
| <hr/> | | | | | |
| RESULT 9 | | | | | |
| AP004782 | 119191 bp DNA linear PRI 11-JUL-2002 | | | | |
| LOCUS | Homo sapiens genomic DNM, chromosome 11 clone:CMB9-103B16, complete sequence. | | | | |
| DEFINITION | | | | | |
| ACCESSION | AP004782 | GI:21728157 | | | |
| VERSION | AP004782.3 | | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Homo sapiens DNA, clone:CMB9-103B16. | | | | |
| ORGANISM | Homo sapiens | | | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| TITLE | Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. | | | | |
| JOURNAL | Homo sapiens genomic DNA | | | | |
| REFERENCE | Published only in Database (2002) | | | | |
| AUTHORS | 2 (bases 1 to 119191) Toyoda,A., Taylor,T.D., Hong-Seog,P., Hattori,M., Ishii,K., Totoki,Y., Watanabe,H. and Sakaki,Y. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (25-FEB-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), genomic Sciences Center (GSC); 1-7-22 Suehiro-hou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattoriegsc.riken.go.jp URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) | | | | |
| COMMENT | On Jul 10, 2002 this sequence version replaced gi:21328197. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| SOURCE | 1..119191 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="11" /map="11q" /clone="CMB9-103B16" | | | | |
| BASE COUNT | 34204 a 26845 c 25557 g 32585 t | | | | |
| ORIGIN | | | | | |
| Query Match | 12.2% Score 45.8 DB 9; Length 119191; | | | | |
| Best Local Similarity | 53.7%; Pred.No.0.34; 82; Indels 0; Gaps 0; | | | | |
| Matches 95; Conservative | 0; Mismatches | | | | |
| OY | 199 GAAGACACGGTCAACAAGAATGTTCATCTCATATGTCGTGATCGTAGTGGAAGACTTTAGACCAA 258 ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ Db 102658 GAAGAAGAAGACGCACGCACGAAGAAGACGCACGCACGCACGAAGAAGACGCACGCACGA 102717 +++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ | | | | |
| OY | 259 GAAAAATTGTGTGCACGAAAACCTTCACCGGAAATTTATTACACGACTCGAGGGGTAA 318 +++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ | | | | |
| Db 102718 | GAAGACGACGACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 102777 +++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ | | | | |
| OY | 319 GATGCCAGGTAAATCCGCGACGATCACGTCAGTCAGATGATGAAGTTGATGACAGATTAA 375 Db 102778 GAAGAACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 102834 +++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ | | | | |
| RESULT 10 | | | | | |
| LOCUS | AF312994 | 228283 bp | DNA | linear | R0D 17-JAN-2002 |
| DEFINITION | Mus musculus chromosome 1 clone MMU, complete sequence. | | | | |
| ACCESSION | AF312994 | | | | |
| VERSION | AF312994.1 | GI:18182288 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Mus musculus. | | | | |
| ORGANISM | Mus musculus | | | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| TITLE | 1 (bases 1 to 228283) Genomic sequence from mouse chr. 1 | | | | |
| JOURNAL | Rump,A., Hess,J., Galoczky,P., Wirth,T. and Rosenthal,A. Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 228283) | | | | |
| AUTHORS | Rump,A., Hess,J., Galoczky,P., Wirth,T. and Rosenthal,A. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (12-OCT-2000) Genome Analysis, Institute of Molecular | | | | |

[illegible]

| | RESULT 11 | | | |
|------------|-------------------------|------------|-------|-----------------------------|
| AF216973 | AF216973 | 20796 bp | DNA | linear |
| LOCUS | Drosophila melanogaster | 1(2)01289 | gene, | complete cds, alternatively |
| DEFINITION | spliced products. | | | |
| ACCESSION | AF216973 | | | |
| VERSION | AF216973.1 | GI:6984064 | | |
| KEYWORDS | | | | |
| SOURCE | | | | |
| ORGANISM | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| FEATURES | | | | |
| source | | | | |
| gene | | | | |
| mrna | | | | |
| mrna | | | | |
| CDS | | | | |

| | a | c | g | t | others |
|------------|-------|-------|-------|-------|--------|
| BASE COUNT | 45406 | 36877 | 36482 | 46377 | 1920 |
| ORIGIN | | | | | |

| | | | |
|----|-------|---|-------|
| QY | 141 | TGATCTTGGAGATGCATAAAAAACATGTTTAAAGACGGAGTACTGCTACATCGTTTTTGA | 200 |
| | | | |
| Db | 43762 | TAATGATGGGGACGACGATAGCGGAGATGACAAATGAGAGACCAAGGACCGAGGATACGA | 43703 |
| QY | 201 | AGACACGGTCACACAGATGTTTACTACAAATGTCGTGGATGGTAAAGATTAAACCAAGA | 260 |
| | | | |
| Db | 43702 | AGGAAGATAAATTAATTAATGACGACGAAGAATGACAGTGAAGACGACGACGAGATGACGACAAAGA | 43643 |
| QY | 261 | AAAAATTTGTTGTCGACGAAAACCTTCACGSGAAAAATTAATTGACAGACTCGGAGGCTTAAGA | 320 |
| | | | |
| Db | 43642 | CAATTAATGATTAACGACGAGGATGACGAAGAAGATTAATTAAGGACGCCACCAATATAACGA | 43583 |
| QY | 321 | TGCACGGTAAATCGCGCAGGTACAGGTGACGAGTCCGATCAAACTGATGAAGATTA | 374 |
| | | | |
| Db | 43582 | AGGAAGATTAATTAATGAAGACGACACAGAGCGATGAGGATTAATAACGACCAAGATGA | 43529 |

| | |
|------------|---|
| RESULT | 15 |
| LOCUS | AC008340/C |
| DEFINITION | 181771 bp DNA linear INV 06-SEP-2001 Drosophila melanogaster, chromosome 2R, region 42D-42E, BAC clone BACRO/JJ20, complete sequence. |
| ACCESSION | AC008340 |
| VERSION | AC008340.5 GI:15451508 |
| KEYWORDS | HTG. |
| SOURCE | Drosophila melanogaster. |
| ORGANISM | Drosophila melanogaster. Neukaryaota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Eukaryaota; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila. 1 (bases 1 to 181771) |
| REFERENCE | Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., AUTHORS |

Celniker, S. E., Adams, M. D., Krommlinger, B., Tyler, D., Wan, K. H.,
Holt, R. A., Evans, A. G., Gocayne, J. D., Ananthakrishnan, P. G.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K. Y.,
Carlson, J. W., Center, A., Chapple, M., Davenport, L. B., Dietz, S. M.,
Podoson, K., Dorey, V., Dou, L. F., Doyle, C., Dreese, D., Farrant, D.,
Fertig, S., Frishe, E., Galle, R. F., Gary, N. S., George, R. A.,
Gonzalez, M., Houch, J., Hoskins, R. A., Hostin, D., Howard, T. J.,
Ildberg, C. C., Jaitani, M., Kruse, D., Li, P., Matzel, B., Moshrefi, A.,
Nichtnisch, T. C., Moy, M., Murphy, B., Neilson, C., Nelson, R. A., Nunn, J.,

| | |
|-----------|---|
| TITLE | Sequencing of Drosophila chromosome 2R, region 42D-42E |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 181771) |
| AUTHORS | Celniker, S.E., Abghyian, A., Arcaina, T.T., Baxter, E., Blazer, R.G., Drosthoff, C., Chame, M., Chavez, C., Chow, M., Ciesiolka, J., |

TITLE Direct Submission
JOURNAL Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Sep 6, 2001 this sequence version replaced gl:13374651.
Sequence submitted by:

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

| FEATURES | source |
|---|---------------------------------|
| Location/Qualifiers | 1..181771 |
| /organism="Drosophila melanogaster" | |
| /strain="y; cn by sp" | |
| /db_xref="taxon:7227" | |
| /chromosome="2R" | |
| /map="42D-42E" | |
| /clone="BACR07J20 (D918)" | |
| /clone_1lb="RPCT-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in PBace3.6)" | |
| BASE COUNT | 50143 a 40177 c 40373 g 51078 t |
| ORIGIN | |

| Query Match | 12.1% | Score 45.2 | DB 3 | Length 181771 |
|-----------------------|---|-----------------------------------|----------------|---------------|
| Best Local Similarity | 49.6% | Pred. Nc 0.49 | | |
| Matches 116 | Conservative | 0 | Mismatches 118 | Indels 0 |
| | | | | Gaps 0 |
| QY 141 | TCATCTTCGACGAATGCAAAAAA | CATGTTTCAGACGAGTACTGCTACATCGTTTGA | 200 | |
| | | | | |
| Db 24403 | TATATGATGGGAGACGAGATGACGAGATGACATATGGAGCGACAMGACGAGGATTAACA | | 24344 | |
| QY 201 | AGACAGGTCACACACAGGATGTTATACATGTCGTTGATGGAGAGTTGACCACAA | | 260 | |
| | | | | |
| Db 24343 | AGACAGTATATATATATGACGACGACAAATGACATGATGAAGACGACGAGATGACGAAGAA | | 24284 | |
| QY 261 | AAAAATTGTTGTCGACGAAAACCTTCACGGAATTTT | TTTGACGAACTGGGAGGGTAAAGA | 320 | |
| | | | | |
| Db 24283 | CAATTAATGATTAACGACGAGAGATGACACAGGAAGTATATATATAGGACGACGAGATTAACA | | 24224 | |
| QY 321 | TGCAGGTATATCGCGACAGGTACAGGTACAGGTCAGATAGATGAAGTTGATGAAGATTAA | | 374 | |
| | | | | |
| Db 24223 | AGAAAGTATATATGAAAGACGACAAAGGACGATGAGATGAAGAAACGACGAAGATGA | | 24170 | |

Search completed: February 25, 2003, 01:54:28
 Job time : 2953 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:11:05 ; Search time 38 Seconds

(without alignments)
2629.946 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 674
Sequence: 1 atgaagttattcttgcatttc.....atgaagtgtgaagatttaa 375

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p.model -DEV-xlh
-O/cgnt2.1/USPTO_spool/US09936737/runat_14022003_100557_1379/app_query.fasta_1.519
-DB-A_Geneseq_101002 -OPMT-fastan -SUFFIX-frag -MIMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS-bits -STRAT=1 -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST=45 -DOCATIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=ptc -NORM-ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09936737_@CGN_1_1_22_@runat_14022003_100557_1379 -NCP=6 -ICPU=3
-NO_XLTPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|----------|---------------------|
| 1 | 573 | 85.0 | 103 | AA994746 | Platelet binding I |
| 2 | 88.5 | 13.1 | 147 | AA922950 | leech antiplatelet |
| 3 | 82.5 | 12.2 | 368 | AB869867 | Drosophila melanog |
| 4 | 82.5 | 12.2 | 412 | AA668010 | S. frugiperda immu |
| 5 | 79.5 | 11.8 | 150 | AA95228 | Mouse soluble BMP |
| 6 | 79.5 | 11.8 | 150 | AA96200 | Soluble bone morph |
| 7 | 79.5 | 11.8 | 200 | AA95234 | Mouse incomplete B |
| 8 | 79.5 | 11.8 | 1038 | AA95224 | Mouse BMP receptor |
| 9 | 79.5 | 11.8 | 1038 | AA96198 | Bone morphogenetic |
| 10 | 79.5 | 11.8 | 1038 | AA96247 | Mouse BMP receptor |
| 11 | 75.5 | 11.2 | 323 | AA958349 | Arabidopsis thalia |
| 12 | 75.5 | 11.2 | 331 | AA958348 | Arabidopsis thalia |
| 13 | 75.5 | 11.2 | 336 | AA958347 | Arabidopsis thalia |
| 14 | 75.5 | 11.2 | 418 | AA914260 | Mouse ice-4 protel |
| 15 | 75.5 | 11.2 | 419 | AA950109 | Murine caspase-12 |
| 16 | 75.5 | 11.2 | 419 | AA914259 | Mouse ice-4 protel |
| 17 | 75.5 | 11.2 | 419 | AA929108 | Murine caspase-12 |
| 18 | 75.5 | 11.2 | 432 | AA96769 | Human Interleukin- |
| 19 | 75.5 | 11.2 | 432 | AA98464 | Murine Ich-3. Mus |
| 20 | 75.5 | 11.2 | 719 | AB927652 | Streptococcus poly |
| 21 | 74.5 | 11.1 | 482 | AB950229 | Human transcriptio |
| 22 | 74.5 | 11.1 | 1547 | AA921168 | Human TRIC-12 pro |
| 23 | 74.5 | 11.0 | 248 | AB966151 | Drosophila melanog |
| 24 | 74 | 11.0 | 577 | AA930586 | Arabidopsis thalia |
| 25 | 74 | 11.0 | 578 | AA930585 | Arabidopsis thalia |
| 26 | 74 | 11.0 | 612 | AA930584 | Arabidopsis thalia |
| 27 | 74 | 11.0 | 1481 | AA903740 | rchd528 gene produ |
| 28 | 74 | 11.4 | 1481 | AA989299 | Amino acid sequenc |
| 29 | 74 | 11.4 | 1481 | AA989299 | Human rchd528 gene |
| 30 | 74 | 11.4 | 1481 | AA919626 | Amino acid sequenc |
| 31 | 74 | 11.4 | 1481 | AA978508 | Richd528 amino acid |
| 32 | 74 | 11.4 | 1481 | AA968447 | Amino acid sequenc |
| 33 | 73.5 | 10.9 | 562 | AA935750 | Endochitinase Ctr52 |
| 34 | 73.5 | 10.9 | 228 | AA944001 | Drosophila melanog |
| 35 | 73.5 | 10.9 | 1154 | AB966151 | Drosophila melanog |
| 36 | 72.5 | 10.8 | 660 | AB966151 | Drosophila melanog |
| 37 | 72.5 | 10.8 | 976 | AA976012 | Human poly(ADP-rib |
| 38 | 72.5 | 10.8 | 976 | AA976012 | Human poly(ADP-rib |
| 39 | 72 | 11.1 | 114 | AA934450 | Zea mays protein f |
| 40 | 72 | 10.7 | 143 | AA920404 | Arabidopsis thalia |
| 41 | 72 | 10.7 | 145 | AA920403 | Arabidopsis thalia |
| 42 | 72 | 10.7 | 587 | AA910554 | S. cerevisiae aspa |
| 43 | 72 | 10.7 | 1092 | AA979135 | Human protein SEQ |
| 44 | 72 | 10.7 | 1092 | AA940100 | Human polypeptide |
| 45 | 72 | 10.7 | 1094 | AA980119 | Human protein SEQ |

ALIGNMENTS

RESULT 1
AA994746
ID AA994746 standard; Protein; 103 AA.

AC AA994746;

DE 29-JAN-2001 (first entry)

Platelet binding inhibitor protein Saratin amino acid sequence.

Saratin: medicinal leech; thromboembolic disease; intraocular lens;

collagen-dependent platelet adhesion inhibition;

posterior capsule opacification.

Hirudo medicinalis.

OS XX
PN WO200056885-A1.
XX
PD 28-SEP-2000.

US-09-936-737A-1 (1-375) x AAR22950 (1-147)

51 ----- 5

QY

QY 52 ACTACTTCTTCAGAGAA-----0

Db 59 Thrilethralaglyasn glyasp cystpserlylsargprogly"trpplsleuProasp /

Db 79 AsnLeuLeuThrLysThrGluPheThrSerValAspGluCysArgLysMetCysGluGlu

QY 226 TACAAT 231

RESULT 3

XX
XX
APP69867:

XX Developmental biology; cell signalling; insecticide; biocontrol

XX
OS *Protophila melanogaster*.

XX
PD 27-SEP-2001.

XX
PR 23-MAR-2000; 2000US-191637P.

AA
PA
NY
(PEKE) PE CORP NY.

XX
DR
WPT: 2001-656860/75.

AA New isolated nucleic acid detection reagent for detecting 1000 or
PT more and for elucidating cell signalling and cell

Enclosure: SEQ ID NO 36393; 21pp + Sequence Listing; English.

Alignment Scores: 1.08 Length: 368

US-09-936-737A-1 (1-375) x ABB69867 (1-368)

```

Db      71 AlaSerGlyLeuProIlnSerAlaSerSerGluAspLeuSerGlnSerLeuSer----- 88

```

Db 89 -----GlutyrThraspalaaspGluserValserAlaProThrGluPheLeuAla 105

Db 106 GlupheuseralavalmetleulysasPTYrLyslysala-----leulysiyr 122

Db 123 CysLysLeuIleuGlnTyrGluProaspAsnAlaThrAlaLysGluPheIyrIleu 142

```

143 170-----LeuAspLysLeuArgAlaValAlaThrSerSerAspSerAspGlu 158
      ::      ::      |||
      |||:::

```

[illegible][illegible]

RESULT 4

XX
XX
AAW68010:

DT 27-APR-1999 (first entry)
XX

XX Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;
KW

XX
05 Spodoptera frugiperda.

| | |
|----------|--------------|
| FT | Domain |
| En | key |
| 111..114 | |
| /vector | "EFAP motif" |

135 138 /note= "putative nuclear localization signal

```

Oy      268 GTTGACGACAAACTCTCAGCGAAAAATTATTGACACTGCGAGGTAAAGATGCAGGT
Db      177 ---AsnAspTherphenylmethasn-----ThrsrAlaGluGlnTyrspspsrsp
Oy      328 AATGCGGACGAGTACAGTGCAGTACAGATGATGATGATGAT 372
Db      194 GluGluAspAspAspGluAspGluGluAspGluAspGluAspAsp 208

RESULT 5
AAR95228
ID AAR95228 standard; Protein: 150 AA.
XX
AC AAR95228;
XX
DT 31-DEC-1996 (first entry)
XX
DE Mouse soluble BMP receptor kinase protein-3 (BRK-3).
XX
KW BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
XX BMP type I receptor kinase; BMP receptor.
XX
OS Mus sp.
XX
PN W09614579-A1.
XX
PD 17-MAY-1996.
XX
PF 30-OCT-1995; 95WO-US14027.
XX
PR 05-JUN-1995; 95US-0462467.
XX PR 04-NOV-1994; 94US-0334178.
XX
PA (PROC ) PROCTER & GAMBLE CO.
XX
PI Rosenbaum JS;
XX
DR WPI: 1996-251887/25.
XX DR N-PADB; AAT28024.
XX
PT Assays for bone morphogenetic protein activities - using complex of
PT BMP type I receptor kinase protein and BMP receptor kinase protein
XX BRK-3
XX
PS Disclosure: Page 63; 101pp; English.
XX
CC Mouse soluble bone morphogenetic protein (BMP) type II receptor
CC kinase protein-3 (BRK-3) (AAR95228) lacks the regions of the full-
CC length receptor (AAR95224) not required for BMP binding, and is
CC the product of a truncated cDNA sequence (AAT28023). A BMP receptor
CC kinase protein complex formed of full-length, truncated, incomplete
CC or soluble BRK-3 and full-length, incomplete or soluble BMP type I
CC receptor kinase proteins (see also AAR95222-27 and AAR95229-34) is
CC useful for screening cpds. for BMP receptor affinity or for
CC determining the concentration of a BMP receptor ligand in a clinical
CC sample. The complex can be expressed by host cells co-transfected
CC with vectors carrying the appropriate DNA sequences (see also
CC AAT28018-30).
XX
XX
SQ Sequence 150 AA:
XX
XX
Alignment Scores:
Pred. NO.:
Score: 1.84 Length: 150
Percent Similarity: 79.50 Matches: 27
Best Local Similarity: 40.46% Conservative: 26
Query Match: 20.61% Mismatches: 35
DB: 11.80% Indels: 43
Gaps: 4
US-09-936-737A-1 (1-375) x AAR95228 (1-150)
Oy      10 TTCTGATTTCCTTCCTTCCTGCGCAAGCTGATCATCTCACTACTCTTACAGACAA 69
||| ::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 9 PheargValProtrpleuLeutrPalValleuLeuValSerThrAlaAlaSerGln 28
 QY 70 CGTGAAGATGTTGGACGTTTACCGCAAGAAATATACAGACTTCGATAAATCTTT 129
 Db 29 AsngInGln-----ArgLeuCysAlaPheLysAspProtyr 40
 QY 130 AAGAACTCTGTGACTGTACGCAATGCAGAAAACA-----TGT 168
 Db 41 GlnGlnAspLeuGlyLeuGlyLeuSerArgIleSerHisGlnAsnGlyThrIleLeuGly 60
 QY 169 TTCAGACGAGACTGCTGACATCGTTTGA-----GACACGGTC 210
 Db 61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
 QY 211 AACAAGAAATGTTACTACATGCGTTCGTAAGAGTTAGACCAAGAAAATTTGTT 270
 Db 81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGlnCysHisTyrGlnGlnCysVal 100
 QY 271 GTC----- 273
 Db 101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
 QY 274 -----GACGAAAACCTTCACGCAAAATTA 297
 Db 121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131
 RESULT 6
 AAR96200 standard; Protein; 150 AA.
 ID AAR96200:
 AC AAR96200:
 XX 22-AUG-1996 (first entry)
 DE Soluble bone morphogenetic protein receptor kinase-3 fragment.
 XX
 KW Mouse; bone morphogenetic protein receptor kinase-3; soluble;
 KW fibroblast; cysteine box; COS; CHO; cell culture; antibody;
 KW diagnostic; bone disorder; osteogenic; drug screening;
 KW bone morphogenetic protein-agonist; reporter gene;
 KW bone morphogenetic protein-antagonist; hormone-responsive element.
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 116..123
 FT /note="Cysteine box"
 XX
 PN W09614412-A2.
 PD 17-MAY-1996.
 XX
 PF 30-OCT-1995; 95WO-US14085.
 PR 04-NOV-1994; 94US-0334179.
 XX
 PA (PROC) PROCTER & GAMBLE CO.
 PI Nohno T, Rosenbaum JS;
 DR WPI, 1996-251762/25.
 DR N-PSDB; AAT27227.
 XX
 PT Isolated bone morphogenic protein receptor kinase protein - used to
 PT determine if a test cpd. is capable of binding to, or is
 PT (ant)agonist of BMP receptor kinase protein transcription
 XX
 XX Claim 4: Page 60; 87pp; English.
 CC The sequence represents a soluble fragment (extracellular domain)
 CC of mouse bone morphogenetic protein (BMP) receptor type-II kinase-3
 CC (BRK-3), which induces cellular differentiation in response to BMP.
 CC A gene encoding full-length BRK-3 (AAR96198) may be isolated from

CC NIH3T3 mouse embryo fibroblast cDNA using PCR. The soluble fragment
 CC binds BMP, and has all regions not required for BMP binding
 CC deleted. Like other transforming growth factor-beta superfamily
 CC receptors, the extracellular domain of BRK-3 contains a single
 CC cysteine box. The BRK-3 receptor gene may be inserted in a vector
 CC and expressed in a CHO or COS cell culture. The receptor and
 CC antibodies against it may be used in diagnostic assays for BMP
 CC disorders, or in therapy to bind or scavenge BMPs. In addition,
 CC expression of the BRK-3 gene along with a reporter gene under the
 CC control of a hormone-responsive element in a cell culture may be
 CC used to screen compounds for BRK-agonist or -antagonist activity,
 CC by monitoring reporter gene expression.
 CC
 SQ Sequence 150 AA:
 Alignment Scores:
 Pred. No.: 1.84 Length: 150
 Score: 79.50 Matches: 27
 Percent Similarity: 40.46% Conservative: 26
 Best Local Similarity: 20.61% Mismatches: 35
 Query Match: 11.80% Indels: 43
 DB: Gaps: 4
 US-09-936-737A-1 (1-375) x AAR96200 (1-150)
 QY 10 TTCTGATTCCTCCCTTCGCGACGTTGCTGATCTCACTCTTCAGAGAA 69
 Db 9 PheargValProtrpleuLeutrPalValleuLeuValSerThrAlaAlaSerGln 28
 QY 70 CGTGAAGATGTTGGACGTTTACCGCAAGAAATATACAGACTTCGATAAATCTTT 129
 Db 29 AsngInGln-----ArgLeuCysAlaPheLysAspProtyr 40
 QY 130 AACAAGATGTTACTACATGCGTTCGTAAGAGTTAGACCAAGAAAATTTGTT 168
 Db 41 GlnGlnAspLeuGlyLeuGlyLeuSerArgIleSerHisGlnAsnGlyThrIleLeuGly 60
 QY 169 TTCAGACGAGACTGCTGACATCGTTTGA-----GACACGGTC 210
 Db 61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
 QY 211 AACAAGAAATGTTACTACATGCGTTCGTAAGAGTTAGACCAAGAAAATTTGTT 270
 Db 81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGlnCysHisTyrGlnGlnCysVal 100
 QY 271 GTC----- 273
 Db 101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
 QY 274 -----GACGAAAACCTTCACGCAAAATTA 297
 Db 121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131
 RESULT 7
 AAR95234 standard; Protein; 200 AA.
 ID AAR95234:
 AC AAR95234:
 XX 31-DEC-1996 (first entry)
 DE Mouse incomplete BMP receptor kinase protein-3 (BRK-3).
 XX
 KW BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
 KW BMP type I receptor kinase; BMP receptor.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Domain 151..172
 FT /label="Transmembrane_domain"
 XX
 PN W09614579-A1.

```

XX 17-MAY-1996.
PD 30-OCT-1995; 95MO-US14027.
XX
XX 05-JUN-1995; 95US-0462467.
XX 04-NOV-1994; 94US-0334178.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Rosenbaum JS;
XX
XX WPI: 1996-251887/25.
XX N-PSDB; AAT28030.
XX
XX Assays for bone morphogenetic protein activities - using complex of
XX BMP type I receptor kinase protein and BMP receptor kinase protein
XX BRK-3
XX
XX Disclosure; Page 80; 101pp; English.
XX
XX Mouse incomplete bone morphogenetic protein (BMP) type II receptor
XX kinase protein-3 (BRK-3) (AAR95234) corresponds to the extracellular,
XX transmembrane and intracellular juxtamembrane region of full-length
XX BRK-3 (AAR95224) but is incapable of signalling due to deletion of the
XX intracellular kinase domain. It is the product of a truncated cDNA
XX sequence (AAT28030). A BMP receptor kinase protein complex formed of
XX full-length, truncated, incomplete or soluble BRK-3 and full-length,
XX incomplete or soluble BMP type I receptor kinase proteins (see also
XX AAR95222-33) is useful for screening cpds. for BMP receptor affinity
XX or for determining the concentration of a BMP receptor ligand in a
XX clinical sample. The complex can be expressed by host cells
XX co-transfected with vectors carrying the appropriate DNA sequences
XX (see also AAT28018-30).
XX
XX Sequence 200 AA:
XX
XX Alignment Scores:
XX Pred. No.: 1.99 Length: 200
XX Score: 79.50 Matches: 27
XX Percent Similarity: 40.46% Conservative: 26
XX Best Local Similarity: 20.61% Mismatches: 35
XX Query Match: 11.80% Indels: 43
XX DB: 17 Gaps: 4
XX
XX US-09-936-737a-1 (1-375) x AAR95234 (1-200)
XX
XX 10 TTCTTGATTTCCTTCCTTCCTCGCAAGCTTGATCTCACTACTCTTCAGAGAA 69
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 9 PheArgValProTrpLeuLeuTrpValLeuValSerThrThrAlaAlaSerGln 28
XX
XX 70 CGTGAAGATTGTTGGACGTTTACCGCAAGAAATATACAGACTTCGATAATCTTTT 129
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 29 AsnGlnGlu-----ArgLeuGlySalPheLysAspProTyr 40
XX
XX 130 AAGAAGTCCTGATCTTGACGAATGCAGAAAACAA-----TCG 168
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 41 GlnGlnAspLeuGlyLeuGlyLeuSerArgLeuSerHisGlnGlnGlyThrLeuGly 60
XX
XX 169 TTCACAGCGGAGTACTGCTACATCGCTTTTGA-----GACACGGTC 210
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspLeuAsnLeuVal 80
XX
XX 211 AACACGAATGTTACTACATATGCTGATGAGTGTAAGAGTTAGACCAAGAAATTTGTT 270
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGlnCysHisTyrGlnGlnCysVal 100
XX
XX 271 GTC----- 273
XX
XX 101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
XX
XX 274 -----GACGAAACCTTCACGGAATAATTAAT 297
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 121 AspleuGlyAsnValAsnPheThrGluAsnPhe 131
RESULT 8
AAR95224
ID AAR95224 standard; Protein; 1038 AA.
XX
XX AAR95224;
XX
XX 31-DEC-1996 (first entry)
XX
XX Mouse BMP receptor kinase protein-3 (BRK-3).
XX
XX BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
XX BMP type I receptor kinase; BMP receptor.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX FT 151..172
XX FT Domain /label=Transmembrane_domain
XX
XX WO9614579-A1.
XX
XX 17-MAY-1996.
XX
XX 30-OCT-1995; 95MO-US14027.
XX
XX 05-JUN-1995; 95US-0462467.
XX 04-NOV-1994; 94US-0334178.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Rosenbaum JS;
XX
XX WPI: 1996-251887/25.
XX N-PSDB; AAT28020.
XX
XX Assays for bone morphogenetic protein activities - using complex of
XX BMP type I receptor kinase protein and BMP receptor kinase protein
XX BRK-3
XX
XX Claim 2; Page 61-62; 101pp; English.
XX
XX Full-length mouse bone morphogenetic protein (BMP) type II receptor
XX kinase protein-3 (BRK-3) (AAR95224) is a receptor capable of binding
XX BMP and transducing a signal initiated by the binding. Its amino
XX acid sequence was deduced from a cDNA clone (AAT28020) obtd. from
XX NIH3T3 fibroblasts. A BMP receptor kinase protein complex formed
XX of full-length, truncated, incomplete or soluble BRK-3 and full-
XX length, incomplete or soluble BMP type I receptor kinase proteins
XX (see also AAR95222-23 and AAR95225-34) is useful for screening cpds.
XX for BMP receptor affinity or for determining the concentration of a
XX BMP receptor ligand in a clinical sample. The complex can be
XX expressed by host cells co-transfected with vectors carrying the
XX appropriate DNA sequences (see also AAT28018-30).
XX
XX Sequence 1038 AA:
XX
XX Alignment Scores:
XX Pred. No.: 3.19 Length: 1038
XX Score: 79.50 Matches: 27
XX Percent Similarity: 40.46% Conservative: 26
XX Best Local Similarity: 20.61% Mismatches: 35
XX Query Match: 11.80% Indels: 43
XX DB: 17 Gaps: 4
XX
XX US-09-936-737a-1 (1-375) x AAR95224 (1-1038)
XX
XX 10 TTCTTGATTTCCTTCCTTCCTCGCAAGCTTGATCTCACTACTCTTCAGAGAA 69
XX ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 9 PheArgValProTrpLeuLeuTrpValLeuValSerThrThrAlaAlaSerGln 28
XX
XX 70 CGTGAAGATTGTTGGACGTTTACCGCAAGAAATATATACAGACTTCGATAATCTTTT 129

```

```

Db      29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
Qy      130 AAGAAGTCTGTGATCTTGACGAAATGCAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyIleuSerArgIleSerHisGlnAsnGlyThrIleLeuCys 60
Qy      169 TTCAGACGAGAGTACTGATCATGCTTTTGA-----GACACGGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
Qy      211 AACAGGAATGTTACTACATGTCGTTGAGGAGGAGTTAACCAAGAAAATTGTT 270
Db      81 LysGlnIleCysTrpSerHisIleGlyAspProGlnIleCysHisTyrGlnIleCysVal 100
Qy      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArpPheCysCysSerThr 120
Qy      274 -----GACGAAAACCTCAGGAAAATTAT 297
Db      121 AspLeuCysAsnValAsnPheThrGluAsnPhe 131

RESULT 9
ID      AAR96198 standard; Protein; 1038 AA.
XX      AAR96198:
XX      AC      AAR96198:
XX      DT      22-AUG-1996 (first entry)
XX      DE      Bone morphogenetic protein type-II receptor kinase-3.
XX      KW      Mouse; bone morphogenetic protein receptor kinase-3; fibroblast;
KW      cysteine box; protein-kinase domain; COS; CHO; antibody;
KW      cell culture; diagnostic; bone disorder; osteogenic;
KW      bone morphogenetic protein-agonist; drug screening; reporter gene;
KW      bone morphogenetic protein-antagonist; hormone-responsive element.
XX      OS      Mus musculus.
XX      FH      Key
FH      Region      Location/Qualifiers
FT      Modified-site      53..57 /note= "Soluble BRK-3 fragment (AAR96199, claim 4)"
FT      Modified-site      110..112 /note= "N-glycosylation site"
FT      Peptide      116..123 /note= "N-glycosylation site"
FT      Peptide      151..172 /note= "Cysteine box"
FT      Domain      173..1038 /note= "Transmembrane region"
FT      Domain      200..504 /note= "Intracellular domain"
FT      Domain      505..1038 /note= "Protein-kinase domain"
FT      Region      /note= "Long C-terminal region"
XX      PN      WO9614412-A2.
XX      PD      17-MAY-1996.
XX      PF      30-OCT-1995; 95WO-US14085.
XX      PR      04-NOV-1994; 94US-0334179.
XX      PA      (PROC ) PROCTER & GAMBLE CO.
XX      PI      Nohno T, Rosenbaum JS;
XX      WPI, 1996-251762/25.
XX      DR      N-PSDB; AAT27225.

```

```

XX      XX      Isolated bone morphogenic protein receptor kinase protein - used to
PT      PT      determine if a test cpd. is capable of binding to, or is
PT      PT      (ant)agonist of BMP receptor kinase protein transcription
XX      XX      Claim 1; Page 56-59; 87pp; English.
XX      CC      The sequence represents mouse bone morphogenetic protein (BMP)
CC      receptor type-II kinase-3 (BRK-3), which induces cellular
CC      differentiation in response to BMP. A gene encoding BRK-3 may be
CC      isolated from NIH3T3 mouse embryo fibroblast cDNA using PCR. The
CC      sequence of a soluble fragment is given in AAR96200. The
CC      extracellular domain of BRK-3 contains a single cysteine box
CC      located near the transmembrane region. The predicted cytoplasmic
CC      region contains all of the consensus sequences characterizing a
CC      protein-kinase domain with predicted specificity for serine and
CC      threonine residues. The kinase domain is followed by an extremely
CC      long C-terminus, making the intracellular domain much larger than
CC      that of any other receptor in the transforming growth factor-beta
CC      receptor family. The BRK-3 receptor gene may be inserted in a
CC      vector and expressed in a CHO or COS cell culture. The receptor
CC      and antibodies against it may be used in diagnostic assays for BMP
CC      disorders, or in therapy to bind or scavenge BMPs. In addition,
CC      expression of the BRK-3 gene along with a reporter gene under the
CC      control of a hormone-responsive element in a cell culture may be
CC      used to screen compounds for BRK-agonist or -antagonist activity,
CC      by monitoring reporter gene expression.
XX      SQ      Sequence      1038 AA;
XX      XX      Alignment Scores:
XX      Pred. No.:      3,19      Length:      1038
XX      Score:      79.50      Matches:      27
XX      Percent Similarity:      40.46%      Conservative:      26
XX      Best Local Similarity:      20.61%      Mismatches:      35
XX      Query Match:      11.80%      Indels:      4
XX      DB:      17      Gaps:      4
XX      US-09-936-737a-1 (1-375) x AAR96198 (1-1038)
Qy      10 TTCTGATTTCCCTTCCTTGCCCTGCGAAGCTTGCTGATCTCAACTCTTCAGAGAA 69
Db      9 PheArgValProTrpLeuLeuTrpAlaValLeuLeuValSerThrThrAlaIleSerGln 28
Qy      70 CGTGAAGATTGTGTGACGTTTTCAGCGAACAAGAAATATACAGACTTCGATAATCTT 129
Db      29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
Qy      130 AAGAAGTCTGTGATCTTGACGAAATGCAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyIleuSerArgIleSerHisGlnAsnGlyThrIleLeuCys 60
Qy      169 TTCAGACGAGAGTACTGATCATGCTTTTGA-----GACACGGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
Qy      211 AACAGGAATGTTACTACATGTCGTTGAGGAGGAGTTAACCAAGAAAATTGTT 270
Db      81 LysGlnIleCysTrpSerHisIleGlyAspProGlnIleCysHisTyrGlnIleCysVal 100
Qy      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArpPheCysCysSerThr 120
Qy      274 -----GACGAAAACCTCAGGAAAATTAT 297
Db      121 AspLeuCysAsnValAsnPheThrGluAsnPhe 131

RESULT 10
ID      AAR96247
XX      AAR96247:
XX      AC      AAR96247:

```

```

XX 16-FEB-1999 (first entry)
DT
XX
XX
DE Mouse BMP receptor kinase protein (BRK)-3.
XX
XX Bone morphogenetic protein; BMP; BMP receptor kinase; ActRIIB receptor;
KW BRK, receptor ligand; drug.
XX
XX
OS Mus sp.
XX
XX WO9852038-A1.
PN
XX
XX 19-NOV-1998.
PD
XX
XX 13-MAY-1998; 98WO-US09519.
PE
XX 16-MAY-1997; 97US-0046768.
PR
XX
XX (PROC ) PROCTER & GAMBLE CO.
PA
XX
XX Rosenbaum JS;
P1
XX
XX WPI: 1999-009930/01.
DR
XX N-PSDB; AAV71971.
DR
XX
XX Screening method using bone morphogenetic protein receptor complex -
PT which binds to potential drugs, and ActRIIB receptor used in the
PT complex, also host cells transfected with DNA encoding the complex
XX
XX
XX Disclosure; Pages 79-84; 110pp; English.
XS

```

CC This represents a mouse BMP receptor kinase protein (BRR)-3. The mouse
CC BRR-1 and BRR-2 protein sequences (AA086248 and AA086249) can be used in
CC the method of the invention of determining whether a compound can bind
CC to a bone morphogenetic protein (BMP) receptor kinase protein complex.
CC The method comprises allowing a protein in a sample to bind to the
CC complex, where the complex is comprised of (i) a BMP; (ii) a BRR protein
CC (iii) an AcvRIIB receptor. The method can be used to determine the
CC concentration of a BMP receptor ligand in a sample by comparing the
CC binding to a standard curve prepared with known concentrations of BMP
CC ligand. The method can also be used to determine whether a test compound
CC produces a signal on binding to a BMP receptor protein complex. The
CC method is useful for determining whether a ligand, such as a known or
CC putative drug, can bind to and/or activate the receptors.

Sequence 1038 AA;

| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 3, 19 |
| Score: | 79.50 |
| Percent Similarity: | 40.44% |
| Best Local Similarity: | 20.61% |
| Query Match: | 11.80% |
| DB: | 20 |
| Length: | 103 |
| Matches: | 27 |
| Conservative: | 26 |
| Mismatches: | 35 |
| Indels: | 43 |
| Gaps: | 4 |

US-09-936-737A-1 (1-375) x AAW86247 (1-1038)

| | | | |
|----|-----|--|-----|
| QY | 10 | TTCTGATTCCTTCCTTGGCTGCGAAGCTGCGATCTCAACTCTCTTCACAGAA | 69 |
| | 11 | :::: ::: ::: ::: | |
| Db | 9 | PheArgValProTPrPLeuThrPAlaValLeuValSerThrThrAlaLeuSerGln | 28 |
| QY | 70 | CCTGAAGATTGTGGACGCTTTTACGCGAACAGAAATATACAGACTTCGATAACTTTT | 129 |
| | 71 | ::::: ::: | |
| Db | 29 | AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr | 40 |
| QY | 130 | AAGAACTCCTCGATTCTTGACGAATGCAAAAAACA-----TGT | 166 |
| | 131 | ::::: :::: | |
| Db | 41 | GlnGlnAspLeuGlyIleGlyGlnSerArgIleSerHisGlnAsnGlyThrIleLeuGly | 60 |
| QY | 169 | TTCAAGACGCGAGTACTGCTACATCGCTTTTGGAA-----GACACGGCTC | 210 |
| | 170 | :::: | |
| Db | 61 | SerLysGlySerThrThrTyrGlyLeuThrGlnLysSerLysGlyAspIleAsnLeuVal | 80 |

| Qy | 211 | AAACAGAAATGTTACTACATGCGTTGATGCGTGAAGACTTTGACCAAGAAAATTTGTT | 270 |
|-----------|--|--|-----|
| | 211 | AAACAGAAATGTTACTACATGCGTTGATGCGTGAAGACTTTGACCAAGAAAATTTGTT | 270 |
| | 81 | LygGlnGlyCysTrpSerHisIleGlyAspProGlnGlyCysHisTyrGlnGlyCysVal | 100 |
| Qy | 271 | GTC----- | 273 |
| | 101 | ValThrThrProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr | 120 |
| Qy | 274 | -----GACGAAATCTTACGGAATTTAT | 297 |
| | 121 | AspLeuCysAsnValAsnPheThrGluAsnIle | 131 |
| RESULT 11 | | | |
| ID | AA658349 | standard; Protein; 323 AA. | |
| XX | AA658349; | | |
| AC | | | |
| XX | 18-OCT-2000 | (first entry) | |
| DE | | | |
| XX | Arabidopsis thaliana protein fragment SEQ ID NO: 75310. | | |
| XX | | | |
| KW | Protein identification; signal transduction pathway; metabolic pathway; | | |
| KW | hybridisation assay; genetic mapping; gene expression control; promoter; | | |
| KW | termination sequence. | | |
| XX | | | |
| OS | Arabidopsis thaliana. | | |
| XX | EP1033405-A2. | | |
| XX | | | |
| PD | 06-SEP-2000. | | |
| XX | | | |
| PF | 25-FEB-2000; 2000EP-0301439. | | |
| XX | | | |
| PR | 25-FEB-1999; | 99US-0121825. | |
| PR | 05-MAR-1999; | 99US-0123180. | |
| PR | 09-MAR-1999; | 99US-0123548. | |
| PR | 23-MAR-1999; | 99US-0125788. | |
| PR | 23-MAR-1999; | 99US-0126264. | |
| PR | 29-MAR-1999; | 99US-0126785. | |
| PR | 01-APR-1999; | 99US-0127462. | |
| PR | 06-APR-1999; | 99US-0128234. | |
| PR | 08-APR-1999; | 99US-0128714. | |
| PR | 16-APR-1999; | 99US-0129845. | |
| PR | 19-APR-1999; | 99US-0130077. | |
| PR | 21-APR-1999; | 99US-0130449. | |
| PR | 23-APR-1999; | 99US-0130510. | |
| PR | 28-APR-1999; | 99US-0130891. | |
| PR | 30-APR-1999; | 99US-0131449. | |
| PR | 30-APR-1999; | 99US-0132048. | |
| PR | 04-MAY-1999; | 99US-0132407. | |
| PR | 05-MAY-1999; | 99US-0132484. | |
| PR | 06-MAY-1999; | 99US-0132485. | |
| PR | 06-MAY-1999; | 99US-0132486. | |
| PR | 07-MAY-1999; | 99US-0132487. | |
| PR | 11-MAY-1999; | 99US-0132663. | |
| PR | 14-MAY-1999; | 99US-0134256. | |
| PR | 14-MAY-1999; | 99US-0134218. | |
| PR | 14-MAY-1999; | 99US-0134219. | |
| PR | 14-MAY-1999; | 99US-0134370. | |
| PR | 18-MAY-1999; | 99US-0134370. | |
| PR | 18-MAY-1999; | 99US-0134768. | |
| PR | 19-MAY-1999; | 99US-0134941. | |
| PR | 20-MAY-1999; | 99US-0135124. | |
| PR | 21-MAY-1999; | 99US-0135353. | |
| PR | 24-MAY-1999; | 99US-0135629. | |
| PR | 25-MAY-1999; | 99US-0136021. | |
| PR | 27-MAY-1999; | 99US-0136392. | |
| PR | 28-MAY-1999; | 99US-0136782. | |
| PR | 01-JUN-1999; | 99US-0137222. | |
| PR | 03-JUN-1999; | 99US-0137528. | |
| PR | 04-JUN-1999; | 99US-0137502. | |
| PR | 07-JUN-1999; | 99US-0137724. | |

Best Local Similarity: 24.11%
 Query Match: 11.20%
 DB: 21
 Mismatches: 41
 Indels: 51
 Gaps: 7

US-09-936-737a-1 (1-375) x AAG58349 (1-323)

OY 10 TTCTGATTCCTTCCTTCCTTCGCAAGCTTCGATCCTCACTACTTCTTCAGAGAA 69
 Db 29 PheValAsnValPhePheCysLeu-----PheLeuLeuLysThrArgSerSerCin 46
 OY 70 CGTGAA-----GATGTGTGACGCTTTACCCGACAGAGAAATATACAGACTTCGAT 120
 Db 47 MetGluSerLeuLeuGluCysPheAlaIleThrAspGlyLysCysHisProAspCysLeu 66
 OY 121 AAATCTTTAAGAGAGCTCTGATCTTGTGACGATGCAAA----- 159
 Db 67 LysAlaAsnAsnGluGluGluLysPheAspAlaCysGlnSerAlaAlaLeuValAlaVal 86
 OY 160 -----AAACATGCTTCAAG-----ACGAGACTGTC 186
 Db 87 SerLeuIleSerSerAlaValAlaIlePheLysIleAspSerLysTyrrGlnIutySer 106
 OY 187 TNCATGCTTTGAAAGACACGCTCAACAGCA----- 219
 Db 107 ProGlnTyrLeuValAspAsnValGlyLysGluGluValGluGluMetAspGlnPro 126
 OY 220 ---TGTTACTACATGTC----- 234
 Db 127 SerCysGlnTyrThrValGlyAsnLeuLeuSerTyrLeuValGluAsnValTrpThrLys 146
 OY 235 -----GTTGATGCTGAGAGAGTATGACCAAGAAAA-----TTTCTTTCGACAGAAAAAC 282
 Db 147 LysGluValAlaArgGlnArgGluMetAspGlnGlnArgArgGlnPheThrValLysAspCys 166
 OY 283 TTC 285
 Db 167 Phe 167

RESULT 12
 AAG58348
 ID AAG58348 standard; Protein; 331 AA.
 XX
 AC AAG58348;
 XX
 DE 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 75309.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0128234.
 PR 06-APR-1999; 99US-0128714.
 PR 08-APR-1999; 99US-0129845.
 PR 16-APR-1999; 99US-0130077.
 PR 19-APR-1999; 99US-0130449.
 PR 21-APR-1999; 99US-0130510.
 PR 23-APR-1999;

PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131448.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134376.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135124.
 PR 25-MAY-1999; 99US-0135124.
 PR 27-MAY-1999; 99US-0135124.
 PR 28-MAY-1999; 99US-0135124.
 PR 01-JUN-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 04-JUN-1999; 99US-0137528.
 PR 07-JUN-1999; 99US-0137502.
 PR 08-JUN-1999; 99US-0137724.
 PR 10-JUN-1999; 99US-0138094.
 PR 14-JUN-1999; 99US-0138547.
 PR 16-JUN-1999; 99US-0138847.
 PR 16-JUN-1999; 99US-0139119.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139464.
 PR 18-JUN-1999; 99US-0139465.
 PR 18-JUN-1999; 99US-0139466.
 PR 18-JUN-1999; 99US-0139467.
 PR 18-JUN-1999; 99US-0139468.
 PR 18-JUN-1999; 99US-0139469.
 PR 18-JUN-1999; 99US-0139470.
 PR 18-JUN-1999; 99US-0139471.
 PR 18-JUN-1999; 99US-0139472.
 PR 18-JUN-1999; 99US-0139473.
 PR 18-JUN-1999; 99US-0139474.
 PR 18-JUN-1999; 99US-0139475.
 PR 18-JUN-1999; 99US-0139476.
 PR 18-JUN-1999; 99US-0139477.
 PR 18-JUN-1999; 99US-0139478.
 PR 18-JUN-1999; 99US-0139479.
 PR 18-JUN-1999; 99US-0139480.
 PR 18-JUN-1999; 99US-0139481.
 PR 18-JUN-1999; 99US-0139482.
 PR 18-JUN-1999; 99US-0139483.
 PR 18-JUN-1999; 99US-0139484.
 PR 18-JUN-1999; 99US-0139485.
 PR 18-JUN-1999; 99US-0139486.
 PR 18-JUN-1999; 99US-0139487.
 PR 18-JUN-1999; 99US-0139488.
 PR 18-JUN-1999; 99US-0139489.
 PR 18-JUN-1999; 99US-0139490.
 PR 18-JUN-1999; 99US-0139491.
 PR 18-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139493.
 PR 18-JUN-1999; 99US-0139494.
 PR 18-JUN-1999; 99US-0139495.
 PR 18-JUN-1999; 99US-0139496.
 PR 18-JUN-1999; 99US-0139497.
 PR 18-JUN-1999; 99US-0139498.
 PR 18-JUN-1999; 99US-0139499.
 PR 18-JUN-1999; 99US-0139500.
 PR 18-JUN-1999; 99US-0139501.
 PR 18-JUN-1999; 99US-0139502.
 PR 18-JUN-1999; 99US-0139503.
 PR 18-JUN-1999; 99US-0139504.
 PR 18-JUN-1999; 99US-0139505.
 PR 18-JUN-1999; 99US-0139506.
 PR 18-JUN-1999; 99US-0139507.
 PR 18-JUN-1999; 99US-0139508.
 PR 18-JUN-1999; 99US-0139509.
 PR 18-JUN-1999; 99US-0139510.
 PR 18-JUN-1999; 99US-0139511.
 PR 18-JUN-1999; 99US-0139512.
 PR 18-JUN-1999; 99US-0139513.
 PR 18-JUN-1999; 99US-0139514.
 PR 18-JUN-1999; 99US-0139515.
 PR 18-JUN-1999; 99US-0139516.
 PR 18-JUN-1999; 99US-0139517.
 PR 18-JUN-1999; 99US-0139518.
 PR 18-JUN-1999; 99US-0139519.
 PR 18-JUN-1999; 99US-0139520.
 PR 18-JUN-1999; 99US-0139521.
 PR 18-JUN-1999; 99US-0139522.
 PR 18-JUN-1999; 99US-0139523.
 PR 18-JUN-1999; 99US-0139524.
 PR 18-JUN-1999; 99US-0139525.
 PR 18-JUN-1999; 99US-0139526.
 PR 18-JUN-1999; 99US-0139527.
 PR 18-JUN-1999; 99US-0139528.
 PR 18-JUN-1999; 99US-0139529.
 PR 18-JUN-1999; 99US-0139530.
 PR 18-JUN-1999; 99US-0139531.
 PR 18-JUN-1999; 99US-0139532.
 PR 18-JUN-1999; 99US-0139533.
 PR 18-JUN-1999; 99US-0139534.
 PR 18-JUN-1999; 99US-0139535.
 PR 18-JUN-1999; 99US-0139536.
 PR 18-JUN-1999; 99US-0139537.
 PR 18-JUN-1999; 99US-0139538.
 PR 18-JUN-1999; 99US-0139539.
 PR 18-JUN-1999; 99US-0139540.
 PR 18-JUN-1999; 99US-0139541.
 PR 18-JUN-1999; 99US-0139542.
 PR 18-JUN-1999; 99US-0139543.
 PR 18-JUN-1999; 99US-0139544.
 PR 18-JUN-1999; 99US-0139545.
 PR 18-JUN-1999; 99US-0139546.
 PR 18-JUN-1999; 99US-0139547.
 PR 18-JUN-1999; 99US-0139548.
 PR 18-JUN-1999; 99US-0139549.
 PR 18-JUN-1999; 99US-0139550.
 PR 18-JUN-1999; 99US-0139551.
 PR 18-JUN-1999; 99US-0139552.
 PR 18-JUN-1999; 99US-0139553.
 PR 18-JUN-1999; 99US-0139554.
 PR 18-JUN-1999; 99US-0139555.
 PR 18-JUN-1999; 99US-0139556.
 PR 18-JUN-1999; 99US-0139557.
 PR 18-JUN-1999; 99US-0139558.
 PR 18-JUN-1999; 99US-0139559.
 PR 18-JUN-1999; 99US-0139560.
 PR 18-JUN-1999; 99US-0139561.
 PR 18-JUN-1999; 99US-0139562.
 PR 18-JUN-1999; 99US-0139563.
 PR 18-JUN-1999; 99US-0139564.
 PR 18-JUN-1999; 99US-0139565.
 PR 18-JUN-1999; 99US-0139566.
 PR 18-JUN-1999; 99US-0139567.
 PR 18-JUN-1999; 99US-0139568.
 PR 18-JUN-1999; 99US-0139569.
 PR 18-JUN-1999; 99US-0139570.
 PR 18-JUN-1999; 99US-0139571.
 PR 18-JUN-1999; 99US-0139572.
 PR 18-JUN-1999; 99US-0139573.
 PR 18-JUN-1999; 99US-0139574.
 PR 18-JUN-1999; 99US-0139575.
 PR 18-JUN-1999; 99US-0139576.
 PR 18-JUN-1999; 99US-0139577.
 PR 18-JUN-1999; 99US-0139578.
 PR 18-JUN-1999; 99US-0139579.
 PR 18-JUN-1999; 99US-0139580.
 PR 18-JUN-1999; 99US-0139581.
 PR 18-JUN-1999; 99US-0139582.
 PR 18-JUN-1999; 99US-0139583.
 PR 18-JUN-1999; 99US-0139584.
 PR 18-JUN-1999; 99US-0139585.
 PR 18-JUN-1999; 99US-0139586.
 PR 18-JUN-1999; 99US-0139587.
 PR 18-JUN-1999; 99US-0139588.
 PR 18-JUN-1999; 99US-0139589.
 PR 18-JUN-1999; 99US-0139590.
 PR 18-JUN-1999; 99US-0139591.
 PR 18-JUN-1999; 99US-0139592.
 PR 18-JUN-1999; 99US-0139593.
 PR 18-JUN-1999; 99US-0139594.
 PR 18-JUN-1999; 99US-0139595.
 PR 18-JUN-1999; 99US-0139596.
 PR 18-JUN-1999; 99US-0139597.
 PR 18-JUN-1999; 99US-0139598.
 PR 18-JUN-1999; 99US-0139599.
 PR 18-JUN-1999; 99US-0139600.
 PR 18-JUN-1999; 99US-0139601.
 PR 18-JUN-1999; 99US-0139602.
 PR 18-JUN-1999; 99US-0139603.
 PR 18-JUN-1999; 99US-0139604.
 PR 18-JUN-1999; 99US-0139605.
 PR 18-JUN-1999; 99US-0139606.
 PR 18-JUN-1999; 99US-0139607.
 PR 18-JUN-1999; 99US-0139608.
 PR 18-JUN-1999; 99US-0139609.
 PR 18-JUN-1999; 99US-0139610.
 PR 18-JUN-1999; 99US-0139611.
 PR 18-JUN-1999; 99US-0139612.
 PR 18-JUN-1999; 99US-0139613.
 PR 18-JUN-1999; 99US-0139614.
 PR 18-JUN-1999; 99US-0139615.
 PR 18-JUN-1999; 99US-0139616.
 PR 18-JUN-1999; 99US-0139617.
 PR 18-JUN-1999; 99US-0139618.
 PR 18-JUN-1999; 99US-0139619.
 PR 18-JUN-1999; 99US-0139620.
 PR 18-JUN-1999; 99US-0139621.
 PR 18-JUN-1999; 99US-0139622.
 PR 18-JUN-1999; 99US-0139623.
 PR 18-JUN-1999; 99US-0139624.
 PR 18-JUN-1999; 99US-0139625.
 PR 18-JUN-1999; 99US-0139626.
 PR 18-JUN-1999; 99US-0139627.
 PR 18-JUN-1999; 99US-0139628.
 PR 18-JUN-1999; 99US-0139629.
 PR 18-JUN-1999; 99US-0139630.
 PR 18-JUN-1999; 99US-0139631.
 PR 18-JUN-1999; 99US-0139632.
 PR 18-JUN-1999; 99US-0139633.
 PR 18-JUN-1999; 99US-0139634.
 PR 18-JUN-1999; 99US-0139635.
 PR 18-JUN-1999; 99US-0139636.
 PR 18-JUN-1999; 99US-0139637.
 PR 18-JUN-1999; 99US-0139638.
 PR 18-JUN-1999; 99US-0139639.
 PR 18-JUN-1999; 99US-0139640.
 PR 18-JUN-1999; 99US-0139641.
 PR 18-JUN-1999; 99US-0139642.
 PR 18-JUN-1999; 99US-0139643.
 PR 18-JUN-1999; 99US-0139644.
 PR 18-JUN-1999; 99US-0139645.
 PR 18-JUN-1999; 99US-0139646.
 PR 18-JUN-1999; 99US-0139647.
 PR 18-JUN-1999; 99US-0139648.
 PR 18-JUN-1999; 99US-0139649.
 PR 18-JUN-1999; 99US-0139650.
 PR 18-JUN-1999; 99US-0139651.
 PR 18-JUN-1999; 99US-0139652.
 PR 18-JUN-1999; 99US-0139653.
 PR 18-JUN-1999; 99US-0139654.
 PR 18-JUN-1999; 99US-0139655.
 PR 18-JUN-1999; 99US-0139656.
 PR 18-JUN-1999; 99US-0139657.
 PR 18-JUN-1999; 99US-0139658.
 PR 18-JUN-1999; 99US-0139659.
 PR 18-JUN-1999; 99US-0139660.
 PR 18-JUN-1999; 99US-0139661.
 PR 18-JUN-1999; 99US-0139662.
 PR 18-JUN-1999; 99US-0139663.
 PR 18-JUN-1999; 99US-0139664.
 PR 18-JUN-1999; 99US-0139665.
 PR 18-JUN-1999; 99US-0139666.
 PR 18-JUN-1999; 99US-0139667.
 PR 18-JUN-1999; 99US-0139668.
 PR 18-JUN-1999; 99US-0139669.
 PR 18-JUN-1999; 99US-0139670.
 PR 18-JUN-1999; 99US-0139671.
 PR 18-JUN-1999; 99US-0139672.
 PR 18-JUN-1999; 99US-0139673.
 PR 18-JUN-1999; 99US-0139674.
 PR 18-JUN-1999; 99US-0139675.
 PR 18-JUN-1999; 99US-0139676.
 PR 18-JUN-1999; 99US-0139677.
 PR 18-JUN-1999; 99US-0139678.
 PR 18-JUN-1999; 99US-0139679.
 PR 18-JUN-1999; 99US-0139680.
 PR 18-JUN-1999; 99US-0139681.
 PR 18-JUN-1999; 99US-0139682.
 PR 18-JUN-1999; 99US-0139683.
 PR 18-JUN-1999; 99US-0139684.
 PR 18-JUN-1999; 99US-0139685.
 PR 18-JUN-1999; 99US-0139686.
 PR 18-JUN-1999; 99US-0139687.
 PR 18-JUN-1999; 99US-0139688.
 PR 18-JUN-1999; 99US-0139689.
 PR 18-JUN-1999; 99US-0139690.
 PR 18-JUN-1999; 99US-0139691.
 PR 18-JUN-1999; 99US-0139692.
 PR 18-JUN-1999; 99US-0139693.
 PR 18-JUN-1999; 99US-0139694.
 PR 18-JUN-1999; 99US-0139695.
 PR 18-JUN-1999; 99US-0139696.
 PR 18-JUN-1999; 99US-0139697.
 PR 18-JUN-1999; 99US-0139698.
 PR 18-JUN-1999; 99US-0139699.
 PR 18-JUN-1999; 99US-0139700.
 PR 18-JUN-1999; 99US-0139701.
 PR 18-JUN-1999; 99US-0139702.
 PR 18-JUN-1999; 99US-0139703.
 PR 18-JUN-1999; 99US-0139704.
 PR 18-JUN-1999; 99US-0139705.
 PR 18-JUN-1999; 99US-0139706.
 PR 18-JUN-1999; 99US-0139707.
 PR 18-JUN-1999; 99US-0139708.
 PR 18-JUN-1999; 99US-0139709.
 PR 18-JUN-1999; 99US-0139710.
 PR 18-JUN-1999; 99US-0139711.
 PR 18-JUN-1999; 99US-0139712.
 PR 18-JUN-1999; 99US-0139713.
 PR 18-JUN-1999; 99US-0139714.
 PR 18-JUN-1999; 99US-0139715.
 PR 18-JUN-1999; 99US-0139716.
 PR 18-JUN-1999; 99US-0139717.
 PR 18-JUN-1999; 99US-0139718.
 PR 18-JUN-1999; 99US-0139719.
 PR 18-JUN-1999; 99US-0139720.
 PR 18-JUN-1999; 99US-0139721.
 PR 18-JUN-1999; 99US-0139722.
 PR 18-JUN-1999; 99US-0139723.
 PR 18-JUN-1999; 99US-0139724.
 PR 18-JUN-1999; 99US-0139725.
 PR 18-JUN-1999; 99US-0139726.
 PR 18-JUN-1999; 99US-0139727.
 PR 18-JUN-1999; 99US-0139728.
 PR 18-JUN-1999; 99US-0139729.
 PR 18-JUN-1999; 99US-0139730.
 PR 18-JUN-1999; 99US-0139731.
 PR 18-JUN-1999; 99US-0139732.
 PR 18-JUN-1999; 99US-0139733.
 PR 18-JUN-1999; 99US-0139734.
 PR 18-JUN-1999; 99US-0139735.
 PR 18-JUN-1999; 99US-0139736.
 PR 18-JUN-1999; 99US-0139737.
 PR 18-JUN-1999; 99US-0139738.
 PR 18-JUN-1999; 99US-0139739.
 PR 18-JUN-1999; 99US-0139740.
 PR 18-JUN-1999; 99US-0139741.
 PR 18-JUN-1999; 99US-0139742.
 PR 18-JUN-1999; 99US-0139743.
 PR 18-JUN-1999; 99US-0139744.
 PR 18-JUN-1999; 99US-0139745.
 PR 18-JUN-1999; 99US-0139746.
 PR 18-JUN-1999; 99US-0139747.
 PR 18-JUN-1999; 99US-0139748.
 PR 18-JUN-1999; 99US-0139749.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139751.
 PR 18-JUN-1999; 99US-0139752.
 PR 18-JUN-1999; 99US-0139753.
 PR 18-JUN-1999; 99US-0139754.
 PR 18-JUN-1999; 99US-0139755.
 PR 18-JUN-1999; 99US-0139756.
 PR 18-JUN-1999; 99US-0139757.
 PR 18-JUN-1999; 99US-0139758.
 PR 18-JUN-1999; 99US-0139759.
 PR 18-JUN-1999; 99US-0139760.
 PR 18-JUN-1999; 99US-0139761.
 PR 18-JUN-1999; 99US-0139762.
 PR 18-JUN-1999; 99US-0139763.
 PR 18-JUN-1999; 99US-0139764.
 PR 18-JUN-1999; 99US-0139765.
 PR 18-JUN-1999; 99US-0139766.
 PR 18-JUN-1999; 99US-0139767.
 PR 18-JUN-1999; 99US-0139768.
 PR 18-JUN-1999; 99US-0139769.
 PR 18-JUN-1999; 99US-0139770.
 PR 18-JUN-1999; 99US-0139771.
 PR 18-JUN-1999; 99US-0139772.
 PR 18-JUN-1999; 99US-0139773.
 PR 18-JUN-1999; 99US-0139774.
 PR 18-JUN-1999; 99US-0139775.
 PR 18-JUN-1999; 99US-0139776.
 PR 18-JUN-1999; 99US-0139777.
 PR 18-JUN-1999; 99US-0139778.
 PR 18-JUN-1999; 99US-0139779.
 PR 18-JUN-1999; 99US-0139780.
 PR 18-JUN-1999; 99US-0139781.
 PR 18-JUN-1999; 99US-0139782.
 PR 18-JUN-1999; 99US-0139783.
 PR 18-JUN-1999; 99US-0139784.
 PR 18-JUN-1999; 99US-0139785.
 PR 18-JUN-1999; 99US-0139786.
 PR 18-JUN-1999; 99US-0139787.
 PR 18-JUN-1999; 99US-0139788.
 PR 18-JUN-1999; 99US-0139789.
 PR 18-JUN-1999; 99US-0139790.
 PR 18-JUN-1999; 99US-0139791.
 PR 18-JUN-1999; 99US-0139792.
 PR 18-JUN-1999; 99US-0139793.
 PR 18-JUN-1999; 99US-0139794.
 PR

```

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145143.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147418.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148655.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155653.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

```

```

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Alignment Scores:

```

Pred. No.: 6.58 Length: 331
Score: 75.50 Matches: 34
Percent Similarity: 34.75% Conservative: 15
Best Local Similarity: 24.11% Mismatches: 41
Query Match: 11.20% Indels: 51
DB: 21 Gaps: 7

```

US-09-936-737a-1 (1-375) x AAG58348 (1-331)

```

QY 10 TTCTTGATTCCTCTCTGCTTCGCGAAGCTTCGATCACTACTCTTCACAGAA 69
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 37 PheValAsnValPhePheCysLeu-----PheLeuLeuLysThrArgSerSerGln 54
QY 70 CGTGA-----GATTGTTGACGCTTTACGCGACAGAAATATACAGACTTCGAT 120
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 55 MetGlnSerLeuLeuGlnCysPheAlaIleThrAspLysCysHisProAspCysLeu 74
QY 121 AATCTTTTACAGACTCTCTGATCTTCGACGATGCAAA----- 159
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75 LysAlaAsnAsnGlnGlnAspLysThrAspLacCysGlnSerAlaAlaLeuValAlaVal 94
QY 160 -----AAACATGTTTCAAG-----ACGAGTACTCGC 186
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 SerLeuIleSerSerAlaArgValIlePheLysIleAspSerLysTYRnGlnTYRser 114
QY 187 TACATCGTTTTCAGACACGCGTCAACAAGAA----- 219
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 115 ProGlnTYRLeuValAspAsnValGlyLysGlnGluValGlnGluGlnMetAspGlnPro 134
QY 220 ---TGTACTACCAATGTC----- 234
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 135 SerCysGlnTYRThrValGlyAsnLeuLeuSerTYRLeuValGlnAsnValTYRThrLys 154
QY 235 -----GTTGATGCTGAAGATTAGACCAAGAAAA-----TTTGTTCGACGAAAC 282
      |||:|:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 LysGlnValArgGlnArgGlnMetAspGlnGlnArgGlnPheThrValLysAspCys 174
QY 283 TTC 285
      |||
DB 175 Phe 175

```

RESULT 13

AAG58347 standard; Protein: 336 AA.

AAG58347;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 75308.

Query Match: 11.20% Indels: 19
 DB: 21 Gaps: 6
 US-09-936-737a-1 (1-375) x AAB14260 (1-418)
 QY 193 GTTTTGAAGACAGCGTCACAAAGAAATGTTACTACATGTCGTTGATGTAAGAGCTTA 252
 |||||:||||| ||| ||| |||||:|||||:|||||:|||||
 Db 26 ValPheAspAspLeuValGluIuys-----AsnValLeuAsnGlyAspGluLeu 41
 253 -----GACCAGAGAAATTTGTTGTCGAC-----GAAACCTTCACCGAAAT 294
 |||:|||||:|||||:|||||:|||||:|||||
 QY 42 LeuIuysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnValGluAsn 61
 ||||| ||||| ||||| |||||:|||||:|||||
 Db 62 PheLeuGluIuysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
 ||||| ||||| ||||| |||||:|||||:|||||
 QY 346 GAGCAG-----TCAGATGAAGTTGAT 366
 |||||:|||||:|||||:|||||
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAsp 94
 |||||:|||||:|||||:|||||
 RESULT 15
 AAY50109 standard; Protein; 419 AA.
 ID AAY50109
 AC AAY50109;
 XX 21-JAN-2000 (first entry)
 DT
 XX Murine caspase-12 splice variant, caspase-12L.
 DE
 XX Caspase: splice variant; truncated; programmed cell death; apoptosis;
 KW regulation; proteolytic cascade; malignant condition; cancer;
 KW premalignant condition; solid tumour; lymphoma;
 KW chronic lymphocytic leukaemia; prostatic hypertrophy;
 KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
 ds.
 KW
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH Cleavage-site 94..95
 FT Protein 95..419
 FT /note="Caspase-12"
 FT Cleavage-site 318..319
 XX
 PN W09952925-A1.
 XX
 PD 21-OCT-1999.
 XX
 XX 14-APR-1999; 99WO-US08064.
 PF
 XX 16-APR-1998; 98US-0081962.
 PR
 XX (GEHO) GEN HOSPITAL CORP.
 PA
 XX Yuan J, Morishima N;
 PI
 DR WPI; 1999-620369/53.
 DR N-PSDB; AA32645.
 DR
 XX New nucleic acid encoding the short form of caspase-12, used e.g. for
 PT treating tumors -
 PT
 XX
 PS Claim 1; Fig 1; 68pp; English.
 CC
 CC This sequence represents a cDNA encoding murine caspase-12L, a splice
 CC variant of caspase-12. Caspases are a family of proteins involved in
 CC the regulation of apoptosis and are synthesised as proforms which are
 CC activated via cleavage after specific Asp residues. Mammalian cells
 CC express several caspases, and it is thought that these act in a
 CC proteolytic cascade to cause programmed cell death. Nucleic acids
 CC encoding caspase-12S (AA32644) or truncated forms of caspase-12L

CC (AA32646, AA32647) are used for production, recombinantly or in vivo,
 CC of caspase-12 polypeptides which induce programmed cell death. This is
 CC particularly useful for treating (pre)malignant conditions (e.g., solid
 CC tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
 CC hypertrophy, preneoplastic liver foci and resistance to chemotherapy), or
 CC autoimmune diseases. The caspase-12 proteins can also be used to raise
 CC specific antibodies (for example, to determine gene expression and to
 CC screen expression libraries) or as molecular weight markers. Fragments
 CC of caspase-12-encoding nucleic acids can be used as probes to isolate
 CC the caspase-12 gene (and its allelic variants); in fluorescent in situ
 CC hybridisation for chromosomal location of the caspase-12 gene, and for
 CC Northern blotting to determine caspase-12 mRNA expression in tissues.
 SQ Sequence 419 AA;
 Alignment Scores:
 Pred. NO.: 7.04 Length: 419
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 11.20% Indels: 19
 DB: 20 Gaps: 6
 US-09-936-737a-1 (1-375) x AAY50109 (1-419)
 QY 193 GTTTTGAAGACAGCGTCACAAAGAAATGTTACTACATGTCGTTGATGTAAGAGCTTA 252
 |||||:||||| ||| ||| |||||:|||||:|||||:|||||
 Db 26 ValPheAspAspLeuValGluIuys-----AsnValLeuAsnGlyAspGluLeu 41
 253 -----GACCAGAGAAATTTGTTGTCGAC-----GAAACCTTCACCGAAAT 294
 |||:|||||:|||||:|||||:|||||:|||||
 QY 42 LeuIuysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnValGluAsn 61
 ||||| ||||| ||||| |||||:|||||:|||||
 QY 295 TATTTG-----ACAGACTGCGAGGTTAA---GATGCAGTAAATCGCGAGGTACAGCT 345
 ||||| ||||| ||||| |||||:|||||:|||||
 Db 62 PheLeuGluIuysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
 ||||| ||||| ||||| |||||:|||||:|||||
 QY 346 GAGCAG-----TCAGATGAAGTTGAT 366
 |||||:|||||:|||||:|||||
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAsp 94
 |||||:|||||:|||||:|||||

Search completed: February 25, 2003, 02:25:51
 Job time : 41 secs

us-09-936-737a-1.raii

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:24:25 ; Search time 13 Seconds
(without alignments)
1697.476 Million cell updates/sec

| | |
|----------------|---|
| Title: | US-09-936-737A-1 |
| Perfect score: | 674 |
| Sequence: | 1 atgaagtattcttgattc.....atgaagttgatgaagattaa 375 |

| Scoring table: | | |
|----------------|------|-------------|
| BROSUIM62 | | |
| xgapop | 10.0 | xgapext 0.5 |
| ygapop | 10.0 | ygapext 0.5 |
| fgapop | 6.0 | fgapext 7.0 |
| delop | 6.0 | delext 7.0 |

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 320146

| | |
|------------------------|-------------|
| Minimum DB seq length: | 0 |
| Maximum DB seq length: | 20000000000 |

| | |
|------------------|--------------------|
| Post-processing: | Minimum Match 0% |
| | Maximum Match 100% |

Listing first 45 summaries

```

Command line parameters: -dev x86_64
MODEL=frimex_n2p_model -runat 14022003.100558.1424/app_query.fasta_1.5181
-O/-cgn2.1/USPO.spool/USO9363737/ -LOOCPct=0.1 -LOOCPct=0
-DB-issued_patents_Ala -ofmt=fasta -SURFEX=rai -MINMATCH=0.1 -TRANS=human40.cdi
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=USO9363737 -SGCN=1.1.4 -emailat 14022003.100558.1424 -NCPU=6 -ICPU=3
-NO.XLPHY -NO.MMAP -LARGESUBSTRY -NES.SCORES=0 -NART -LONGIOS -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEX=0.5 -FGAPOP=6 -FGAPEX=7
-XGAPOP=10 -YGAPEX=0.5 -DELOP=6 -DELEXT=7

```

```

Database : Issued_Patents_A6
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed
and is derived by analysis of the total score distribution.

```

SUMMARIES

| Result No. | Score | Query | Match length | DB | ID | Description |
|------------|-------|-------|--------------|----|-------------------|-------------------|
| 1 | 88.5 | 13.1 | 147 | 1 | US-08-044-547-1 | Sequence 1, Appl |
| 2 | 82.5 | 12.2 | 412 | 2 | US-08-741-134-2 | Sequence 2, Appl |
| 3 | 80.5 | 11.9 | 126 | 1 | US-08-044-547-3 | Sequence 3, Appl |
| 4 | 79.5 | 11.8 | 150 | 4 | US-08-334-179A-10 | Sequence 10, Appl |
| 5 | 79.5 | 11.8 | 1038 | 4 | US-08-334-179A-8 | Sequence 8, Appl |
| 5 | 75.5 | 11.2 | 172 | 4 | US-09-187-789-16 | Sequence 16, Appl |
| 6 | 75.5 | 11.2 | 172 | 4 | US-09-139-600-11 | Sequence 11, Appl |
| 7 | 73.5 | 11.2 | 418 | 3 | US-08-258-287B-58 | Sequence 58, Appl |
| 8 | 75.5 | 11.2 | 418 | 3 | US-08-368-704C-56 | Sequence 56, Appl |
| 9 | 75.5 | 11.2 | 418 | 3 | US-08-368-704C-55 | Sequence 55, Appl |
| 10 | 75.5 | 11.2 | 419 | 3 | US-08-258-287B-57 | Sequence 57, Appl |
| 11 | 75.5 | 11.2 | 419 | 3 | US-08-368-704C-55 | Sequence 55, Appl |
| 12 | 74 | 11.4 | 1481 | 2 | US-08-616-844-40 | Sequence 40, Appl |

| | | | | | | |
|---|----|------|------|------|---|--------------------|
| C | 13 | 74 | 11.4 | 1481 | 2 | US-08-599-6584-40 |
| C | 14 | 74 | 11.4 | 1481 | 3 | US-08-944-8684-40 |
| C | 15 | 74 | 11.4 | 1481 | 3 | US-08-944-428A-40 |
| C | 16 | 74 | 11.4 | 1481 | 3 | US-08-944-496-40 |
| C | 17 | 73.5 | 10.9 | 562 | 6 | 5258502-2 |
| C | 18 | 72.5 | 10.8 | 976 | 4 | US-09-302-812-4 |
| C | 19 | 72.5 | 10.8 | 976 | 4 | US-09-511-477-4 |
| C | 20 | 72.5 | 10.8 | 976 | 4 | US-09-511-507-4 |
| C | 21 | 72 | 10.7 | 2555 | 3 | US-09-058-489-36 |
| C | 22 | 71.5 | 10.6 | 146 | 2 | US-08-888-497-34 |
| C | 23 | 71.5 | 10.6 | 146 | 5 | US-09-462-230-34 |
| C | 24 | 71.5 | 10.6 | 146 | 5 | US-09-462-230-34 |
| C | 25 | 71 | 10.5 | 129 | 4 | PCT-US94-07926-34 |
| C | 26 | 71 | 10.5 | 129 | 4 | US-09-134-001C-45A |
| C | 27 | 70.5 | 10.5 | 849 | 4 | US-09-454-667-1 |
| C | 28 | 69.5 | 10.3 | 124 | 1 | US-09-157-457-4 |
| C | 29 | 69.5 | 10.3 | 450 | 2 | US-08-170-360-1 |
| C | 30 | 69.5 | 10.3 | 450 | 2 | US-08-712-9398-1 |
| C | 31 | 69.5 | 10.3 | 674 | 4 | US-09-265-930-1 |
| C | 32 | 68.5 | 10.5 | 335 | 4 | US-08-961-083-800 |
| C | 33 | 68.5 | 10.2 | 359 | 2 | US-09-403-768-6 |
| C | 34 | 68 | 10.1 | 425 | 1 | US-08-713-636-2 |
| C | 35 | 68 | 10.1 | 425 | 3 | US-08-700-789A-5 |
| C | 36 | 68 | 10.1 | 425 | 3 | US-09-020-684-5 |
| C | 37 | 68 | 10.1 | 425 | 3 | US-09-020-457-5 |
| C | 38 | 68 | 10.1 | 425 | 3 | US-09-020-685-5 |
| C | 39 | 67.5 | 10.0 | 388 | 4 | US-09-020-683-5 |
| C | 40 | 67.5 | 10.0 | 450 | 4 | US-09-265-630-11 |
| C | 41 | 67.5 | 10.0 | 450 | 4 | US-09-265-630-6 |
| C | 42 | 67 | 9.9 | 124 | 1 | US-09-265-630-13 |
| C | 43 | 67 | 9.9 | 124 | 4 | US-08-170-360-2 |
| C | 44 | 67 | 9.9 | 124 | 4 | US-08-888-497-41 |
| C | 45 | 67 | 9.9 | 124 | 5 | US-09-362-230-41 |
| | | | | | | PCT-US94-07926-4 |

ALIGNMENTS

RESULT 1
US-08-044-547-1
Application US/08044547

SEQUENCE 1, APPLICATION
PATENT NO. 5324715

GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M

APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting a Thrombin-Stimulated Platelet Aggregation

| | |
|----------------------|------|
| TITLE OF INVENTION: | COLL |
| NUMBER OF SEQUENCES: | 12 |

CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc

STREET: P.O. Box ;
STW: Railway;

;;
;;
;;
CITY: Ramsey
STATE: New Jersey

COUNTRY: US
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
SYSTEM: PC-DOS/MS-DOS #1.0
OPERATING: version #1.25

```

SOFTWARE: PatentIn Release #1.0,
CURRENT APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

```

; FILING DATE:
; CLASSIFICATION: 514
; ERROR APPLICATION DATA:

```

PRIOR APPLICATION NUMBER: US 07/594,911
APPLICATION NUMBER: 08-OCT-1990

FILING DATE: 03-OCT-1990
 ATTORNEY/AGENT INFORMATION:

NAME: PARR, RICHARD S.
REGISTRATION NUMBER: 32,586

REFERENCE/DOCKET NUMBER: 1
TELECOMMUNICATION INFORMATION

TELEPHONE: (908) 594-4
TELEFAX: (908) 594-472

TELEFAX: (202) 335-1111

10

OS-08-044-547-1 (1-147)

CURRENT APPLICATION DATA: windows 0.1

STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey

```

? APPLICATION NUMBER: US/08/741,134
? FILING DATE:
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60,007,163
? FILING DATE: 01-NOV-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Deluca, Mark
? REGISTRATION NUMBER: 33,229
? REFERENCE/DOCKET NUMBER: TJU-2090
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-568-3100
? TELEFAX: 215-568-3439
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 412 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US -08-741-134-2

Alignment Scores:
Pred. No.: 0.0899 Length: 412
Score: 82.50 Matches: 32
Percent Similarity: 41.74% Conservative: 16
Best Local Similarity: 27.83% Mismatches: 40
Query Match: 12.24% Gaps: 27
2 5

US-09-936-737A-1 (1-375) x US-08-741-134-2 (1-412)

QY 43 CTGATCTCAACTACTTCTTCAGAGAAGCGTAAGATTGTGGACGTTTAAAGCGAACGA 107
| | | | | : : : | | | | | 
Db 116 LeuValProAlaIysAsnLysArgLysLeuGIuAsnAlaAsnAspAlaThrAlaAsnLys 135
| | | AAmATATCAGACTTCGATAAATCTTTAAGAAAGTCC-----TCGATCTT 147
| | | LysAlaIysProAspLysLysAlaGlyLysAsnSerAlaProAlaAlaGluSerAspSer 155
QY 148 GACCAATGCAAAAAACAAGTTTCAAAGACGGAGTGCTCATCGCTTTTGAAGACACG 207
| | | | | : : : | | | | | 
Db 156 AspAspAspAsp-----GluAspGln 162
QY 208 GTCAACAAGAAATGTTACTACATATCGTTGATGTCGAGAGTTAGACCAAGAAAATT 267
| | | | | : : | | | | | : : : : 
Db 163 LeuGlnLys-----PheLeuAspGlyGluAspLysPheAspGlu--- 176
QY 268 GTTGTGCAGAAATCTCACGAAATATTGTGCACACTGCCAGGGTAAAGATGCAGGT 327
| | | | | : : | | | | | : : : : 
Db 177 ---AsnAspGluSerPheLysMetAsn-----ThSerAlaGlnLysAspSerAsp 193
QY 328 AATGCGGCGGTACGAGTGCAGAGATCGATGAAGATTGATGACMAAT 372
Db 194 GIUGLuAspAspAspGluAspGLuAspGLuAspGLuAspAspAsp 208

RESULT 3
US-08-044-547-3
; Sequence 3, Application US/08044547
; Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-Stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSER: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
```

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/044,547
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/594,917
 FILING DATE: 09-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Parr, Richard S.
 REGISTRATION NUMBER: 32,586
 REFERENCE/DOCKET NUMBER: 18053
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-4958
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 126 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-044-547-3

Alignment Scores:
 Pred. No.: 0.109 Length: 126
 Score: 80.50 Matches: 18
 Percent Similarity: 50.91% Conservative: 10
 Best Local Similarity: 32.73% Mismatches: 24
 Query Match: 11.94% Indels: 3
 Gaps: 1

US-09-936-737a-1 (1-375) x US-08-044-547-3 (1-126)

QY 76 GATTGTTGGACGTTTACGGAAGAAATATACAGACTTCGATTAATCTTTAAGAG 135
 |||||
 Db 45 AspGyStrpSerlySarprGlyTrpLysleuProAspAsnleuThrlGlu 64
 |||||
 QY 136 TCCCTGATCTTGAAGATGCAAAAACATGTTTCAAGACGATAC-----TGC 186
 ::::|
 Db 65 PhehrSerValAspGlySarGlyMetCysGluGluSerAlaValGluProSerCys 84
 ::::|
 QY 187 TACATCGTTTGAAGACAGCGTCACAAAGAAATGTTACTACAT 231
 |||||
 Db 85 TyrIleuGlnIleAsnThrGluThrAsnGluCysTyrArgAsn 99
 |||||

RESULT 4
 US-08-334-179A-10
 ; Sequence 10, Application US/08334179A
 ; Patent No. 6306622
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSENBAUM, JAN S.
 ; APPLICANT: NOHNO, TSUTOMU
 ; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
 ; STREET: 11810 EAST MIAMI RIVER ROAD
 ; CITY: ROSS
 ; STATE: OH
 ; COUNTRY: US
 ; ZIP: 45061
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.30, Version #1.30
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/334,179A
 FILING DATE: 04-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CONSTANJE, BRAHM J.
 REGISTRATION NUMBER: 34,804
 REFERENCE/DOCKET NUMBER: 5473
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 513-627-2858
 TELEFAX: 513-627-0260
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 150 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-334-179A-10

Alignment Scores:
 Pred. No.: 0.152 Length: 150
 Score: 79.50 Matches: 27
 Percent Similarity: 40.46% Conservative: 26
 Best Local Similarity: 20.61% Mismatches: 35
 Query Match: 11.80% Indels: 43
 Gaps: 4

US-09-936-737a-1 (1-375) x US-08-334-179A-10 (1-150)

QY 10 TTCCTGATTTCCCTTCTTCTTCCGGAAGCTTGATGATCTGATCTTTCAGAA 69
 |||||
 Db 9 PheArgValProThrPleuLeuThrAlaValIleuLeuValSerThrThrAlaValSerGln 28
 |||||
 QY 70 CGTGAAGATTGTTGACGTTTACGCGAAGACAGAAATATACAGACTTCGATTAATCTTT 129
 ::::|
 Db 29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
 ::::|
 QY 130 AAGAAGTCTCTGATCTTGAAGATGCAAAAACAA-----TGT 168
 ::::|
 Db 41 GlnGlnAspLeuGlyIleGlyGluSerArgIleSerHisGluAsnGlyThrIleLeuCys 60
 ::::|
 QY 169 TTCAGAGCGAGTACTGCTCATCGTTTGA-----GACACGGTTC 210
 |||||
 Db 61 SerIlyGlySerThrCysTyrGlyLeuThrGluLysSerIlyAspIleAsnLeuVal 80
 |||||
 QY 211 AACAGAGATGTTACTACAAATGCTGATGTCGAGAGTTAGACCAAGAAATTTGTT 270
 ::::|
 Db 81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGluCysHisTyrGlnGluCysVal 100
 ::::|
 QY 271 GTC----- 273
 |||||
 Db 101 ValThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
 |||||
 QY 274 -----GACGAAACTTCACGGAATATAT 297
 ::::|
 Db 121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131
 ::::|

RESULT 5
 US-08-334-179A-8
 ; Sequence 8, Application US/08334179A
 ; Patent No. 6306622
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSENBAUM, JAN S.
 ; APPLICANT: NOHNO, TSUTOMU
 ; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
 ; STREET: 11810 EAST MIAMI RIVER ROAD
 ; CITY: ROSS
 ; STATE: OH
 ; COUNTRY: US
 ; ZIP: 45061
 ; COMPUTER READABLE FORM:

| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 0.276 |
| Score: | 79.50 |
| Percent Similarity: | 40.45% |
| Best Local Similarity: | 20.61% |
| Query Match: | 11.80% |
| DB: | 4 |
| Gaps: | 43 |

00 05-530-15/A-1 (1-375) X US-08-334-179A-8 (1-1038)

[illegible]

RESULT 6
 US-09-187-789-16
 ; Sequence 16 Application US/09187789
 ; Patent No. 6340740
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Alnemri, Emdad S.
 ; APPLICANT: Fernandez-Alnemri, Teresa
 ; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
 ; TITLE OF INVENTION: AND METHODS OF USE
 ; FILE REFERENCE: 480140.434C1
 ; CURRENT APPLICATION NUMBER: US/09/187,789
 ; CURRENT FILING DATE: 1998-11-06

```

; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 172
; TYPE: prt
; ORGANISM: Mus musculus
US-09-187-789-16

```

Alignment Scores:
Pred No.:

| | | | |
|------------------------|--------|---------------|----|
| Record No.: | 0.487 | Length: | 17 |
| Score: | 75.50 | Matches: | 26 |
| Percent Similarity: | 54.79% | Conservative: | 14 |
| Best Local Similarity: | 35.62% | Mismatches: | 14 |
| Query Match: | 11.20% | Indels: | 19 |
| DB: | 4 | Gaps: | 6 |

100-15741 (1-3/5) X US-09-187-789-16 (1-172)

[illegible]

```

RESULT 7
US-09-139-600-11
/ Sequence 11, Application US/09139600
/ Patent No. 6432628
/ GENERAL INFORMATION:
/ APPLICANT: Alnemri, Emed S.
/ APPLICANT: Fernandez-Alnemri, Teresa
/ TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
/ FILE REFERENCE: 460140.434
/ CURRENT APPLICATION NUMBER: US/09/139.600
/ NUMBER OF SEQ ID NOS: 65
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 11
/ LENGTH: 172
/ TYPE: PRT
/ ORGANISM: Mus musculus
US-09-139-600-11

```

Alignment Scores:

| | | | |
|------------------------|--------|---------------|-----|
| Query NO.: | 0.487 | Length: | 172 |
| Score: | 75.50 | Matches: | 16 |
| Percent Similarity: | 54.798 | Conservative: | 24 |
| Best Local Similarity: | 35.628 | Mismatches: | 14 |
| Query Match: | 11.208 | Indels: | 19 |
| DB: | 4 | Gaps: | 6 |

US-09-139-600-11 (1-172), x US-09-139-600-11 (1-375)

| | | | |
|----|-----|--|--------------------------------|
| QY | 193 | GTTTTGAACACCGTCAACAGCAATGTTTACTACATGCTGTTGATGGTGAAGAGTTA | 25 |
| | | ::: | ::: ::: |
| Db | 26 | ValPheaspaspIeuValGluLys----- | AsnValIeuAsnGlyAspIuIeu |
| | | ValLys----- | 41 |
| QY | 253 | -----GACCAGAGAAAATTGTTGTCGAC----- | GAAAACTTCACGGAATAAT |
| | | ::: ::: | 294 |
| Db | 42 | IeuLysIleGlyGluSerPheIleIeuAsnLysAlaGluAsnIeuValGluAsn | 61 |
| | | ::: ::: | |
| QY | 295 | TATTTG-----ACAGACTCGAGAGGTAAA----- | GATCGACGTAATAGCGCGGACGATACAGGT |
| | | ----- | 345 |

Db 62 PheLeuIuIystrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
 Oy 346 GACGAG-----TCAGATGAAGTTGAT 366
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAspAsp 94

RESULT 8

US-08-258-287B-58
 ; Sequence 58, Application US/08258287B
 ; Patent No. 6083735
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuan, Junying
 ; APPLICANT: Miura, Masayuki
 ; TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
 ; NUMBER OF SEQUENCES: 85
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/258,287B
 ; FILING DATE: 10-JUN-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/080,850
 ; FILING DATE: 24-JUN-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bugalsky, Lawrence B.
 ; REGISTRATION NUMBER: 35,086
 ; REFERENCE/DOCKET NUMBER: 0609,3920001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2540
 ; TELEFAX: (202) 371-2540
 ; TELEX: 248636 SSK
 ; INFORMATION FOR SEQ ID NO: 58:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 418 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-258-287B-58

Alignment Scores:
 Pred. No.: 0.639 Length: 418
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 11.20% Indels: 19
 DB: 3 Gaps: 6

US-09-936-737A-1 (1-375) x US-08-258-287B-58 (1-418)

Oy 193 GTTTTGAAGACAGCGTCACAGGAATGTACTACATGTCGTTGANGTGAGAGGTTA 252
 Db 26 ValPheAspAspLeuValGluLys-----AsnValLeuAsnGlyAspGluLeu 41
 Oy 253 -----GACCAAGAAAAATTTGTTGTCGAC-----GAAACTTCACGGAAAAAT 294
 Db 42 LeuLysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnLeuValGluAsn 61
 Oy 295 TATTTG-----ACAGACTGCGAGGTTAAA---GATGCGAGTAATGCGGACAGTACAGGT 345
 Db 62 PheLeuGluLysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81

Oy 346 GACGAG-----TCAGATGAAGTTGAT 366
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAspAsp 94

RESULT 9

US-08-368-704C-56
 ; Sequence 56, Application US/08368704C
 ; Patent No. 6087160
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuan, Junying
 ; APPLICANT: Miura, Masayuki
 ; TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
 ; NUMBER OF SEQUENCES: 95
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/368,704C
 ; FILING DATE: 4-JAN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/258,287
 ; FILING DATE: 10-JUN-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/080,850
 ; FILING DATE: 24-JUN-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bugalsky, Lawrence B.
 ; REGISTRATION NUMBER: 35,086
 ; REFERENCE/DOCKET NUMBER: 0609,3920002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2540
 ; TELEFAX: (202) 371-2540
 ; TELEX: 248636 SSK
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 418 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-368-704C-56

Alignment Scores:
 Pred. No.: 0.639 Length: 418
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 11.20% Indels: 19
 DB: 3 Gaps: 6

US-09-936-737A-1 (1-375) x US-08-368-704C-56 (1-418)

Oy 193 GTTTTGAAGACAGCGTCACAGGAATGTACTACATGTCGTTGANGTGAGAGGTTA 252
 Db 26 ValPheAspAspLeuValGluLys-----AsnValLeuAsnGlyAspGluLeu 41
 Oy 253 -----GACCAAGAAAAATTTGTTGTCGAC-----GAAACTTCACGGAAAAAT 294
 Db 42 LeuLysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnLeuValGluAsn 61
 Oy 295 TATTTG-----ACAGACTGCGAGGTTAAA---GATGCGAGTATGCGGACAGTACAGGT 345


```

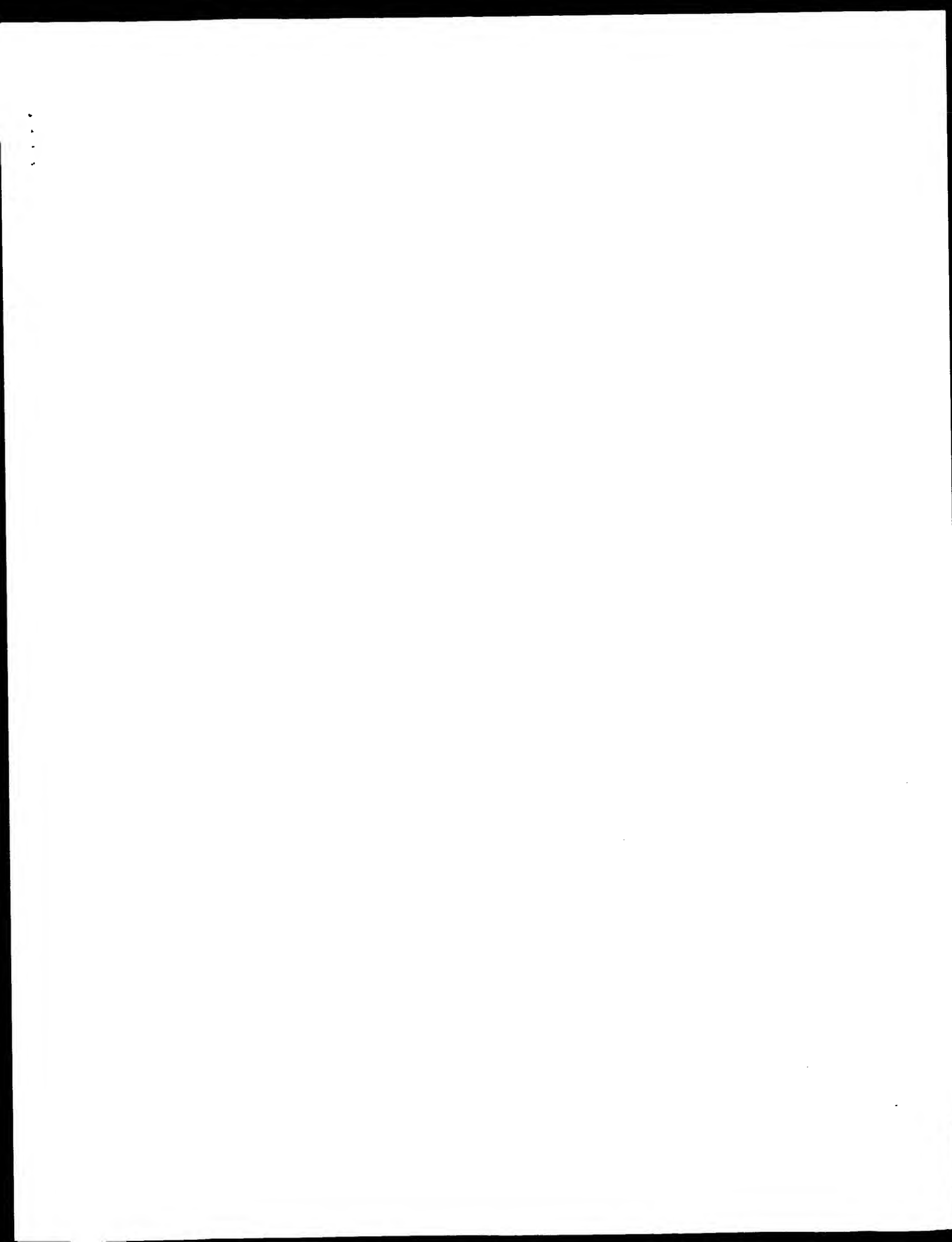
Db      880  -----SerginSerThrProHisGlnIuLys 888
QY      56  GTAGTTGAGATGAGCAG 39
      |||::: |||||
Db      889  ValIleThrGluSerLys 894

Search completed: February 25, 2003, 02:29:51
Job time : 17 secs

```

| Pred. No.: | Alignment Scores: | Length: |
|-------------------------------|-------------------|---------|
| Score: 1.44 | Matches: 1481 | |
| Percent Similarity: 74.00 | Conservative: 43 | |
| Best Local Similarity: 39.04% | Mismatches: 19 | |
| Query Match: 29.45% | Indels: 50 | |
| DB: 11.37% | Gaps: 7 | |

| US-09-936-737A-1 (1-375) | x | US-08-944-423A-40 (1-1481) |
|---|---|----------------------------|
| QY 371 TCTTCAACATTCAGCTGACTGCTCACTGCTGACCGCAATTACCT----- | | 324A |
| Db 764 SerSerSerSerSerSerSerSerSerSerGjProProLeuProLeuProSerValSerGln | | 7833 |
| QY 323 -----GCATCTTTACCTCGCAGCTGTGCAAAATTAATTTCCGTGAG | | 2822 |
| Db 784 SerHisHisLeuPheSerSerSerSerLeuLeuProSerThrArgAlaSerValHisLeuLeuLys | | 8033 |
| QY 281 TTTTCTCGCAGCAACAATTTTCTTGCTCACTCT---TCACCA-----TCACG | | 234 |
| Db 804 SerThrSerAlaSerThrProTyrSerSerSerProSerProLeuProValSerLeu | | 8233 |
| QY 233 ACATGTAGTAACATTCCTTTGTGACCGGTCTTCAAAACGATGACGACTGCTC--- | | 177 |
| Db 824 ThrThrSerThrSerAlaProLeuSerValSerGlnThrThrLeuProGlnSerSerSer | | 8433 |
| QY 176 -----GTCTTGAACAT----- | | 165 |
| Db 844 ThrProValLeuProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet | | 863 |
| QY 164 -----GTTTTTTGCATTCGTCAGATCAGAGACTCTTAAAGATTATGCG | | 117 |
| Db 864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAsp---LeuLys--- | | 879 |
| QY 116 AAGTGTGATATATTCTTGTTGGCGGTAAACGTCACCACTTCACGTTCTTGTGAAGA | | 57 |



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:25:55 ; Search time 31 Seconds

(without alignments)
751.689 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 674
Sequence: 1 atgaagattcttcttgattc.....atgaagttgataagattaa 375

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 156504 segs, 31069816 residues

Total number of hits satisfying chosen parameters: 313008

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09936737/rnat_14022003_100558_1452/app_query.fasta_1.519
-DB=Published Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09936737/cgn_1_1_2_@rnat_14022003_100558_1452
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:

1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description | |
|------------|-------|--------------------|------|------------------|------------------|
| 1 | 75.5 | 11.2 | 172 | US-09-989-903-16 | |
| 2 | 74 | 11.0 | 1178 | US-10-041-856-9 | |
| 3 | 74 | 11.4 | 1481 | US-09-371-900-40 | |
| c | 4 | 74 | 11.4 | 1481 | US-09-924-417-60 |

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|-------|---------------------|
| 5 | 73.5 | 10.9 | 562 | US-09-801-368-82 |
| 6 | 72.5 | 10.8 | 976 | US-08-973-451-4 |
| 7 | 72 | 10.7 | 143 | US-09-867-550-2038 |
| 8 | 71.5 | 10.6 | 156 | US-09-833-790-252 |
| 9 | 70.5 | 10.5 | 578 | US-09-821-839-2 |
| 10 | 70 | 10.4 | 1167 | US-09-815-242-11522 |
| 11 | 69.5 | 10.3 | 674 | US-09-765-272-200 |
| 12 | 68.5 | 10.5 | 72 | US-09-764-660-558 |
| 13 | 68.5 | 10.2 | 206 | US-09-738-626-6537 |
| 14 | 68 | 10.1 | 146 | US-09-993-999-7 |
| 15 | 68 | 10.4 | 556 | US-09-801-368-426 |
| c | 16 | 67.5 | 281 | US-09-895-9134-282 |
| 17 | 67.5 | 10.0 | 1036 | US-09-918-508-4 |
| 18 | 67.5 | 10.0 | 1092 | US-10-135-322-18 |
| 19 | 67 | 9.9 | 322 | US-09-872-523-5 |
| 20 | 66.5 | 9.9 | 2789 | US-09-801-574-57 |
| 21 | 66 | 9.8 | 172 | US-10-087-195-6 |
| 22 | 66 | 9.8 | 172 | US-10-085-572-6 |
| 23 | 66 | 9.8 | 485 | US-10-140-372-11 |
| 24 | 65.5 | 9.7 | 258 | US-09-808-602-110 |
| 25 | 65.5 | 9.7 | 270 | US-09-816-028A-39 |
| 26 | 65 | 9.6 | 374 | US-09-820-004-2 |
| 27 | 65 | 9.6 | 417 | US-09-820-004-4 |
| 28 | 65 | 9.6 | 417 | US-09-820-004-5 |
| 29 | 65 | 9.6 | 417 | US-09-820-004-6 |
| c | 30 | 65 | 661 | US-10-094-214-5 |
| 31 | 65 | 9.6 | 1125 | US-09-974-296-114 |
| 32 | 65 | 10.0 | 2344 | US-09-815-242-12713 |
| c | 33 | 65 | 26926 | US-09-759-508B-2 |
| 34 | 64.5 | 9.6 | 156 | US-09-925-297-511 |
| 35 | 64.5 | 9.6 | 977 | US-09-973-451-2 |
| 36 | 64.5 | 9.6 | 1038 | US-09-908-500A-2 |
| 37 | 64 | 9.5 | 225 | US-10-185-770-4 |
| 38 | 64 | 9.5 | 272 | US-10-125-852-23 |
| 39 | 64 | 9.5 | 292 | US-10-125-852-23 |
| 40 | 64 | 9.5 | 497 | US-09-745-763-166 |
| 41 | 64 | 9.5 | 262 | US-09-416-344A-7 |
| 42 | 64 | 9.5 | 590 | US-10-042-417-52 |
| c | 43 | 63.5 | 376 | US-09-801-368-106 |
| 44 | 63.5 | 9.4 | 571 | US-09-925-301-1031 |
| 45 | 63.5 | 9.4 | 829 | US-09-946-805-8 |

ALIGNMENTS

RESULT 1
US-09-989-903-16
; Sequence 16, Application US/09989903
; Patent No. US20020146804A1
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emdad S.
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434D1
; CURRENT APPLICATION NUMBER: US/09/989,903
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-989-903-16

Alignment Scores:

Pred. No.: 0.84
Score: 75.50
Percent Similarity: 54.79%
Best Local Similarity: 35.62%
Query Match: 11.20%
DB: 10
Length: 172
Matches: 26
Conservative: 14
Mismatch: 14
Indels: 19
Gaps: 6

| | | | |
|----|-----|--|-----|
| QY | 15 | TCCTTCCTTTGCGCGAAGCTTCGATGCTCAACTCTTTCAGAGAACGTCAA--- | 75 |
| | | | |
| | | | |
| Db | 400 | SerialGcysLeuSerAspIleValAlaIatYThrThrSerThHisIstValIstVal | 411 |
| QY | 76 | -----GATGTGGACCTTTTACGCGACACAGAAATATATCA | 111 |
| | | | |
| | | | |
| Db | 420 | IleThrSerAspTrpIlyIleIleSerCysMetLeuPheIlyIstIlystIstArgAsn | 435 |
| QY | 112 | GACTTCGATAATCTTTTAAAGAGTCCTCGATCTTGACGAATGCAAAAAACATGTTTC | 171 |
| | | ::: | |
| | | | |
| Db | 440 | TyrSerAsnProPhePheArgIlyIstYrIleLeuGluIleIleuIlyValProSerHis | 455 |
| QY | 172 | AAGACGGAGTAC---TCCTCATCGCTTTTGAAGACACGGTCACACAGAACTGTACTAC | 222 |
| | | | ::: |
| | | | |
| Db | 460 | IlyThrTrpPheAlaCysPheAlaIstValSerGlnAspThr----- | 472 |
| QY | 229 | AAATGCTGTGATGCTGAAGAGTTAGACCAAGAAATTTTGTTCGACGAA | 279 |
| | | | |
| | | | |
| Db | 473 | -----AspGlyTyrIlyIstPheAsnSerAspArgAlaSerIleAspGlu | 486 |

RESULT 3
 US-09-371-900-40
 Sequence 40, Application US/09371900

| Alignment Scores: | 2.27 | Length: | 1481 |
|------------------------|--------|---------------|------|
| Pred. No.: | 74.00 | Matches: | 43 |
| Score: | 39.04% | Conservative: | 14 |
| Percent Similarity: | 29.45% | Mismatches: | 39 |
| Best Local Similarity: | 11.37% | Indels: | 50 |
| Query Match: | 10 | Gaps: | 7 |

DB: US-09-924-417-60

US-09-936-737A-1 (1-375) x US-09-924-417-60 (1-1481)

324

371 TCTTATCAACTTATCTGATCTGACCTGACCTGTAACCTGCCGCACTTACCT----- 324

DB 764 SerSerSerSerSerSerSerSerSerGlyProProLeuProLeuProSerValSerGln 783

323 -----GCATCTTTACCTCGACGTCGTCAATAAATTTTCCGCGAG 803

DB 784 SerHisHisLeuPheSerSerSerLeuLeuProSerThrArgAlaSerValHisLeuLeuLys 803

321 TTTTCGTCGCAACAATAATTTTCTTGCTTAACCT---TCACCA-----TCACG 234

DB 804 SerThrSerAlaSerThrProThrSerSerSerProSerProLeuProValSerLeu 823

223 ACATGTAGTACATCTTCCTTGACCGCTGTTCAAAAAGATGTAGACAGATCC--- 177

DB 824 ThrThrSerThrSerAlaProLeuSerValSerGlnThrThrLeuProGlnSerSer 843

176 -----GCTCTTGAACAT----- 165

DB 844 ThrProValLeuProArgAlaArgGlnThrProValThrSerPheGlnThrSerThrMet 863

164 -----GTTTTTTCATTCGTGACAGATCAGAGACTTTAAAGAATTATCG 117

DB 864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAsp---LeuLys 879

116 AAGTCGTATATTTCCTTGTCGCGTAAACGCCAACATCTTCACGTTCTTGAAGAA 57

DB 880 -----SerGlnSerThrProHisGlnLeuLys 888

56 GTAGTTGAGATCAGCAG 39

DB 889 ValIleThrGlnSerLys 894

RESULT 5

US-09-801-368-82 Application US/09801368

Sequence 82, Patent No. US20020128250A1

GENERAL INFORMATION:

APPLICANT: Busby, Robert

APPLICANT: Cali, Brian

APPLICANT: Hecht, Peter

APPLICANT: Holtzman, Doug

APPLICANT: Madden, Kevin

APPLICANT: Maxon, Mary

APPLICANT: Milne, Todd

APPLICANT: No. US20020128250A1man, Thea

APPLICANT: Royer, John

APPLICANT: Salama, Sofie

APPLICANT: Sherman, Amir

APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric

TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

FILE REFERENCE: 109272.147

CURRENT APPLICATION NUMBER: US/09/801.368

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: US 09/487,558

PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: US 60/160,587

PRIOR FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 440

SOFTWARE: PatentIn version 3.0

SEQ ID NO 82

LENGTH: 562
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-09-801-368-82

Alignment Scores:

Pred. No.: 1.99
 Score: 73.50
 Percent Similarity: 27
 Best Local Similarity: 39.81%
 Query Match: 25.00%
 DB: 10.91%
 Gaps: 1

US-09-936-737a-1 (1-375) x US-09-801-368-82 (1-562)

QY 29 GCCTCCGACGCTTGTGATCTCACTCTCTTCAAGAAACGTTGAAGATGTTGGACGT 88
 Db 311 AIsaSerGlnThrAlaThrThrValAlaThrSerLysThrSerAlaAlaSerThrSer 330
 QY 89 TTTACGCGAAGAAATATACAGACTTCGATTAATCTTTAAGAAGTCCTGATCTTG 148
 Db 331 SerAlaSerThrSerSerAlaSerThrSerGlnLysLysThrThrGlnSerThrSer 350
 QY 149 ACGAATGCAAAAAACATGTTTCAAGACGAGTACGTACATGCTTTTGAAGACACGG 208
 Db 351 ThrGlnSerLysSerLysValThrLeuSerProThrAlaSerSerAlaLleLysThrSer 370
 QY 209 TCACAGAGAAATGTTACTACATGCTGATGCTGAAGAGTTAACCAGAAATTTG 268
 Db 371 IleThrGlnThrThrLysThrLeuThrSerThrLysThr-----LysSerSer 387
 QY 269 TTGTCGACGAAACTTCACGCGAAATTTATTTGACAGACTCGGAGGTAAGATGACGTA 328
 Db 388 LeuGlyThrThrThrThrLysThrLysThrLysThrLysThrLysThrLysThr 407
 QY 329 ATGCGGAGGTACAGGTGACGAGT 352
 Db 408 ThrLeuSerSerGlnIleThrSer 415

RESULT 6

US-09-973-451-4
 Sequence 4, Application US/09973451
 Patent No. US20020132328A1
 GENERAL INFORMATION:
 APPLICANT: JACOBSON, Myron K.
 APPLICANT: JACOBSON, Elaine L.
 APPLICANT: AM, Jean-Christophe
 APPLICANT: LIN, Winston
 TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOPHROLASE
 TITLE OF INVENTION: (PARG) ENZYMES,
 TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
 FILE REFERENCE: NIAD 201
 CURRENT APPLICATION NUMBER: US/09/973,451
 PRIOR APPLICATION NUMBER: 2001-10-09
 PRIOR FILING DATE: 1999-04-30
 PRIOR APPLICATION NUMBER: 60/083,768
 PRIOR FILING DATE: 1998-05-01
 NUMBER OF SEQ ID NOS: 38
 SEQ ID NO 4
 LENGTH: 976
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 US-09-973-451-4

Alignment Scores:

Pred. No.: 3.04
 Score: 72.50
 Percent Similarity: 28
 Best Local Similarity: 38.89%
 Query Match: 25.93%
 DB: 10.76%
 Length: 976
 Matches: 28
 Conservative: 14
 Mismatches: 35
 Indels: 31

DB: 10 Gaps: 5

US-09-936-737a-1 (1-375) x US-09-973-451-4 (1-976)

QY 49 TCACTACTCTTCACAAAGACGTGAGATGTTGGACGTT-----TAC 93
 Db 197 SerAlaSerThrSerLysThrLysThrLysThrLysThrLysThrLysThrLys 216
 QY 94 GCGAAGCAAAATATACAGACTTCGATTAATCTTTAAGAAGTCTCTGATCTTGACGAA 153
 Db 217 AlaAlaLysGlnThrThrGlnLysAlaThrGlnLysThrLysThrLysThrLys 236
 QY 154 TCAGAAAAACATGTTTCAAGACGAGTACGTCATGCTTTGAAGACAGGTGAC 213
 Db 237 CysSerLysSerCysHisProGlyLysAspCys----- 247
 QY 214 AAGAAATGTTTCAATGCTGATGTTGAAGACTTACACCAAGAAATTTGTTGTC 273
 Db 248 AlaSerCysGlnGlnAspGlnIleAsp-----ValVal 258
 QY 274 GACGAAACCTTCACGCGAAATTTATTTGACAGACTCGAGGTTAAGATGACGATATGCG 333
 Db 259 -----ProLysSerProLysSerSerPalaLysSerGlnAsp----- 270
 QY 334 GCAGTACAGGTGACGATGACAT 357
 Db 271 ValGlyThrGlySerLysAsnAsp 278

RESULT 7

US-09-867-550-2038
 Sequence 2038, Application US/09867550
 Patent No. US20020082206A1
 GENERAL INFORMATION:
 APPLICANT: Leach, Martin D.
 APPLICANT: Mehraban, Ebad
 APPLICANT: Conley, Pamela
 APPLICANT: Law, Debbie
 APPLICANT: Topper, James
 TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a
 TITLE OF INVENTION: Thereby
 FILE REFERENCE: 21402-013 (Cura-313)
 CURRENT APPLICATION NUMBER: US/09/867,550
 PRIOR FILING DATE: 2001-09-20
 PRIOR APPLICATION NUMBER: USSN 60/208,427
 PRIOR FILING DATE: 2000-05-30
 NUMBER OF SEQ ID NOS: 2125
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2038
 LENGTH: 143
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-867-550-2038

Alignment Scores:

Pred. No.: 2.05
 Score: 72.00
 Percent Similarity: 42.40%
 Best Local Similarity: 23.20%
 Query Match: 10.68%
 DB: 10
 Length: 143
 Matches: 29
 Conservative: 24
 Mismatches: 60
 Indels: 12
 Gaps: 5

US-09-936-737a-1 (1-375) x US-09-867-550-2038 (1-143)

QY 10 TTTCTGATTCCTCTCTCTTGCCTCGAAGCTTGCATCTCACTACTTTTGAAGAA 69
 Db 15 TTTCTGATTCCTCTCTCTTGCCTCGAAGCTTGCATCTCACTACTTTTGAAGAA 69
 QY 70 CGTGAAGATGTTGGACGCTTTTACCGAAGCAAAATATACAGACTTCGATTAATCTTT 129
 Db 34 -----CysLeuHisThrPheCysLys-----ThrCysLleValGlnHisPhe 47
 QY 130 AAGAGTCTCTGATCTTGAAGCAAAATATGTTTCAAGCGAGTACTGCTAC 189
 Db 34 -----CysLeuHisThrPheCysLys-----ThrCysLleValGlnHisPhe 47

| | | | |
|--|-----|--|-----|
| Db | 122 | gIyArGgIaInTtYrGcIyArGgIyArGserGlyGInValArGaSpGpUuTtYrArGInAsp | 141 |
| RESULT 9 | | | |
| US-09-821-839-2 | | | |
| Sequence 2, Application US/09821839 | | | |
| Patent No. US20020129407A1 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Ma, Hong | | | |
| TITLE OF INVENTION: Plant Gene Required for Male Meiosis | | | |
| FILE REFERENCE: Psu-0020 | | | |
| CURRENT APPLICATION NUMBER: US/09/821,839 | | | |
| CURRENT FILING DATE: 2001-03-29 | | | |
| PRIOR APPLICATION NUMBER: 60/193,523 | | | |
| PRIOR FILING DATE: 2000-03-31 | | | |
| NUMBER OF SEQ ID NOS: 7 | | | |
| SOFTWARE: PatentIn version 3.1 | | | |
| SEQ ID NO 2 | | | |
| LENGTH: 578 | | | |
| TYPE: PRT | | | |
| ORGANISM: Arabidopsis thaliana | | | |
| US-09-821-839-2 | | | |
| Alignment Scores: | | | |
| Pred. No.: 4.51 Length: 578 | | | |
| Score: 70.50 Matches: 24 | | | |
| Percent Similarity: 37.61% Conservative: 20 | | | |
| Best Local Similarity: 20.51% Mismatches: 52 | | | |
| Query Match: 10.46% Indels: 21 | | | |
| Db: 10 Gaps: 4 | | | |
| US-09-936-737A-1 (1-375) x US-09-821-839-2 (1-578) | | | |
| Qy | 58 | TCCTTCAGAGACGTCGATGATTGTGGACGTTTACGCGACAGAAAAATATACAGACTTC | 117 |
| Db | 265 | SetSerGluGlnArGserGluIeTyrSerGlnTyrSerAspPheAspTyrSerAspTyr | 284 |
| Qy | 118 | GATAAATCTT-----TTTAAAGAAAGTCCTGTATCTTACGGAATGCAAAAAACA----- | 165 |
| Db | 285 | ThrProSerIlePheAspSerGlySerGluPheSerGluTyrSerSerAspSer | 304 |
| Qy | 166 | -----TGTTCAGACGAGTACTGCTACATGCTT | 195 |
| Db | 305 | ProIleSerHisSerTyrSerIleuTyrIleuGlnPheGlyGlnIlePheCysArgSerThr | 324 |
| Qy | 196 | TTTGAAGACACGCTCAACAAGAAATGTTACTACATGTC-----GTT | 237 |
| Db | 325 | IleProAsnAspPheGlySerSerCysGluGluIleHisSerGluLeuLeuArgPhe | 344 |
| Qy | 238 | GATGGTGGAAGAGTTAGACCAAGAAAAATTTGTTGTGCGAGAAACTTCACGGAAT-- | 294 |
| Db | 345 | AspAspGluGlnValGluIleGluIleGluSerTyrIleuArgLeuArgGluArgGluArgSerHisAla | 364 |
| Qy | 295 | TATTTGACACAGCTCGAGGGTTAAAGATGACAGGTATGCGGACAGTACAGT | 345 |
| Db | 365 | TyrMetArgSpCysAlaIlySalatYrCysSerArgMetAspAsnThrIly | 381 |
| RESULT 10 | | | |
| US-09-815-242-11522 | | | |
| Sequence 11522, Application US/09815242 | | | |
| Patent No. US20020061569A1 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Haselbeck, Robert | | | |
| APPLICANT: Ohlsen, Karl L. | | | |
| APPLICANT: Zyskind, Judith W. | | | |
| APPLICANT: Wall, Daniel | | | |
| APPLICANT: Trawick, John D. | | | |
| APPLICANT: Carr, Grant J. | | | |
| APPLICANT: Yamamoto, Robert T. | | | |
| TITLE OF INVENTION: Identification of Essential Genes in | | | |
| Prokaryotes | | | |
| FILE REFERENCE: ELITRA 011A | | | |

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001

TYPE: PRT
ORGANISM: Homo sapiens

TYPE: PRT
ORGANISM: Homo sapiens

Alignment Scores:
Pred. No.: 4.36 Length: 72
Score: 68.50 Matches: 18
Percent Similarity: 51.67% Conservative: 13
Best Local Similarity: 30.00% Mismatches: 18
Query Match: 10.52% Indels: 11
DB: 10 Gaps: 2

US-09-936-737a-1 (1-375) x US-09-764-860-558 (1-72)

QY 324 TGCATCTTACCTCGAGTC-----TGCAAAATATTTTC 289
Db 6 CysvalpheuInAlaLysThrThrThrasnLysproHisCysArgIleGluasn 25

QY 288 CGTAAATTTTC-----GTGCAACAATAATTTTCTTGCTTACCTCCACCATC 238
Db 26 ArgserCysphecysrlyglYleuArgArgvalPheTyrllelleuLeuMetIleleu 45

QY 237 AACACATTTGATTAACATCTTGCACCGCTCTTCAAAAACATGATGACATCTC 178
Db 46 AsnInmetleuLeuThrPheleuValAspCysArgTyserThrAsnIleAlaValleu 65

RESULT 13
US-09-738-626-6537
; Sequence 6537, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6537
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6537

Alignment Scores:
Pred. No.: 5.82 Length: 206
Score: 68.50 Matches: 25
Percent Similarity: 54.32% Conservative: 19
Best Local Similarity: 30.86% Mismatches: 26
Query Match: 10.16% Indels: 11
DB: 9 Gaps: 3

US-09-936-737a-1 (1-375) x US-09-738-626-6537 (1-206)

QY 134 AGTCCTGTGATCTTGAAGAAATGTTTCAAGACGAGTACTGCTACATCG 193
Db 129 SerProgluInleAlaPheAsnLeuArgLysPheLeuSerIleProProSerIleSer 148

QY 194 TTTTGAAGACAGCGTCAACAGGATGTTACTACATGCTTGTATGGGAAGATTAG 253
Db 149 AspleuProThrArgLys---ArgAspValleuThrleuLeuHisLeuGlyLysSer--- 166

QY 254 ACCAAGAAAAATTTGTTTCGACGAAACTTCACGGAAATTTATTTGACAGACTCGAG 313
Db 167 -----AsnAlaAspIleAlaIaGluLeuLysSerThrValSer 180

QY 314 GTAAGATCGAGTA-----ATGCGCAGGTACAGTGCAGATGATGAAGTTG 364
Db 181 ValLysLysSerValSerArgLeuMetGlnArgPheGlnValSerSerArgLeuLeu 200

QY 365 ARG 367
Db 201 Val 201

RESULT 14
US-09-993-999-7
; Sequence 7, Application US/09993999
; Patent No. US20020110891A1
; GENERAL INFORMATION:
; APPLICANT: Ho, I-Cheng
; APPLICANT: Arm, Jonathan P.
; APPLICANT: Austen, K. Frank
; APPLICANT: Glincher, Laurie H.
; TITLE OF INVENTION: Phospholipase A2 Group Preferentially
; FILE REFERENCE: HU1-046
; CURRENT APPLICATION NUMBER: US/09/993, 999
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 60/246, 316
; PRIOR FILING DATE: 2000-11-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-993-999-7

Alignment Scores:
Pred. No.: 6.06 Length: 146
Score: 68.00 Matches: 15
Percent Similarity: 40.00% Conservative: 9
Best Local Similarity: 25.00% Mismatches: 18
Query Match: 10.09% Indels: 18
DB: 10 Gaps: 2

US-09-936-737a-1 (1-375) x US-09-993-999-7 (1-146)

QY 142 GATCTTGAAGAAATGCAAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTTGA 201
Db 62 AspleuAspArgCys-----CysGlnThrHisAspHisCysTy----- 74

QY 202 GACACGGTCAACACAGGATGTTACTACAAATGCTTGATGGTGAAGATTGACCAAGA 261
Db 75 -----SerGlnAlaLysLysLeuGluSerCys 83

QY 262 AAATTTGTTGCGAGAAACTTCACGGAAATTTATTTGACAGACTCGGAGGTAAGAT 321
Db 84 LysPheleuIleAspAsnProIlyThrAsnThrTySerTyserCysSerIleu 103

RESULT 15
US-09-801-368-426
; Sequence 426, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Rhea
; APPLICANT: Royer, John

```

; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 426
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-426
```

```

Alignment Scores:
Pred. No.:      8.75      Length:      556
Score:          68.00      Matches:      27
Percent Similarity: 50.67%      Conservative: 11
Best Local Similarity: 36.00%      Mismatches: 35
Query Match:    10.45%      Indels:      2
DB:             10      Gaps:      1
```

US-09-936-737a-1 (1-375) x US-09-801-368-426 (1-556)

```

QY  371 TCTTCATCAACTTCATGCTGACGTCACCTGCGCATTTACCTGCAATCTTACCC 312
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB  225 SerSerThrThrSerSerThrThrSerSerThrThrSerSerThrThrSerThr 244
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY  311 TCGCAGTCTGCAAAATTAATTTCCGTCGGAAGTTTCGTCGACACAAATTTCTTGCT 252
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB  245 SerSerThrThrSer-----SerThrThrSerSerThrThrSerIlePheSerValThr 262
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY  251 AACTCTTCGCCATCAACGACATGTAGTAACATTCCTGTTGACCGTGCTTCACAAACG 192
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB  263 SerSerSerSerSerIleThrLeuSerSerSerGluHisThrValAlaSpSerArgThr 282
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY  191 ATGTAGCAGTACTCGCTCTTGAACATGTTTTTTTGCATTCGCTCA 147
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB  283 SerSerProSerSerThrLeuValProValSerSerSerSerSer 297
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

Search completed: February 25, 2003, 02:31:05
Job time : 33 secs

Run on: February 24, 2003, 20:43:39 ; search time 202 seconds
(without alignments)

Title: US-09-936-737A-1

Sequence: 1 atgaagtattcttgatttc.....atgaagllgaclygaagaccua 3'

Scoring table: IDENTITY_NDC
Gapex 10.0 : Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

| | | |
|------------------|---------------|------|
| Post-processing: | Minimum Match | 0% |
| | Maximum Match | 100% |

Listing first 45 summaries

```
Database : N_Geneseq_101002:*
1. /sids2/accddata/geneseq/geneseqn-emb1/NA1980.DAT:*
```

| | |
|-----|--|
| 1: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.* |
| 2: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.* |
| 3: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.* |
| 4: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.* |
| 5: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.* |
| 6: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.* |
| 7: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.* |
| 8: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.* |
| 9: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.* |
| 10: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.* |
| 11: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.* |
| 12: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.* |
| 13: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.* |
| 14: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.* |
| 15: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.* |
| 16: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.* |
| 17: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.* |
| 18: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.* |
| 19: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.* |
| 20: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.* |
| 21: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.* |
| 22: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.* |
| 23: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.* |
| 24: | /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|-----|----------------------------------|---------------------------|
| 1 | 375 | 100 | 0 | 375 | 21 | AAA28180 Platelet binding |
| 2 | 45.8 | 12.2 | 49999 | 20 | AAZ23891 Murine LOBO genome | |
| 3 | 45.8 | 12.2 | 49999 | 20 | AAZ23896 Murine LOBO genome | |
| 4 | 45.2 | 12.1 | 15935 | 23 | ABU06114 Drosophila melanogaster | |
| 5 | 44.4 | 11.8 | 2438 | 24 | ABO34916 Human ovarian antigen | |
| 6 | 41 | 10.9 | 372 | 23 | AAE54552 DNA encoding novel | |
| 7 | 40.2 | 10.7 | 3211 | 18 | AAE89346 Human p160 CDNA 1 | |
| 8 | 40.2 | 10.7 | 3901 | 18 | AAE89345 Human p160 CDNA 1 | |
| 9 | 39.6 | 10.6 | 766 | 22 | ABA49279 Human breast cells | |

| | | | | | | | |
|---|----|------|------|---------|----|-----------|---------------------|
| C | 10 | 39.6 | 10.6 | 766 | 22 | ABA67193 | Human foetal liver |
| C | 11 | 39.6 | 10.6 | 766 | 22 | ABA34288 | Probe #12754 for g |
| C | 12 | 39.6 | 10.6 | 766 | 22 | AAL12114 | Probe #12047 for g |
| C | 13 | 39.6 | 10.6 | 766 | 22 | AAI47409 | Probe #16095 used t |
| C | 14 | 39.6 | 10.6 | 766 | 22 | AAI07812 | Probe #7803 used t |
| C | 15 | 39.6 | 10.6 | 1944 | 22 | ABAA44123 | Human breast cell |
| C | 16 | 39.6 | 10.6 | 1944 | 22 | ABBA54575 | Human foetal liver |
| C | 17 | 39.6 | 10.6 | 1944 | 22 | ABBA24359 | Probe #2825 for ge |
| C | 18 | 39.6 | 10.6 | 1944 | 22 | AAI132874 | Probe #2807 for ge |
| C | 19 | 39.6 | 10.6 | 1944 | 22 | AAI34221 | Probe #2917 used t |
| C | 20 | 39.6 | 10.6 | 1944 | 21 | AAAI02791 | Probe #2782 used t |
| C | 21 | 39.6 | 10.6 | 3489 | 21 | AAA30290 | Kaposi's sarcoma-a |
| C | 22 | 39.6 | 10.6 | 3489 | 22 | AAAF82901 | Nucleotide sequenc |
| C | 23 | 39.6 | 10.6 | 3207 | 20 | ABBA3487 | Kaposi's sarcoma-a |
| C | 24 | 39.6 | 10.6 | 3207 | 20 | AAAV3805 | KSHV LVR DNA (nucl |
| C | 25 | 39.6 | 10.6 | 137507 | 19 | AAVI9941 | KSHV Long unique c |
| C | 26 | 39.4 | 10.5 | 354 | 23 | AAAS0677 | DNA encoding novel |
| C | 27 | 39.4 | 10.5 | 654 | 23 | AAAS9553 | DNA encoding novel |
| C | 28 | 39.4 | 10.5 | 654 | 23 | AAAS71153 | DNA encoding novel |
| C | 29 | 39.4 | 10.5 | 654 | 23 | AAAS75467 | DNA encoding novel |
| C | 30 | 39.2 | 10.5 | 234 | 23 | AAAS75460 | DNA encoding novel |
| C | 31 | 39 | 10.4 | 400 | 23 | AAAS68580 | Arabidopsis thalia |
| C | 32 | 39 | 10.4 | 963 | 23 | AAAC7605 | Arbidopsis thalia |
| C | 33 | 38.6 | 10.3 | 1448 | 21 | AAAG7038 | DNA encoding novel |
| C | 34 | 38.4 | 10.2 | 1072 | 23 | AAAS90738 | Listeria monocytol |
| C | 35 | 37.8 | 10.1 | 2944528 | 24 | ABBA03041 | DNA encoding novel |
| C | 36 | 37.6 | 10.0 | 1622 | 23 | AAAS64269 | Plasmodium falcipa |
| C | 37 | 37.4 | 10.0 | 2658 | 21 | AAAT02123 | Shrimp white spot |
| C | 38 | 37.4 | 10.0 | 3543 | 22 | AAAH62819 | Shrimp white spot |
| C | 39 | 37.4 | 10.0 | 305107 | 22 | AAAH62869 | Human breast cell |
| C | 40 | 37.2 | 9.9 | 315 | 22 | ABAA51452 | Human breast cell |
| C | 41 | 37.2 | 9.9 | 315 | 22 | ABAA6847 | Human foetal liver |
| C | 42 | 37.2 | 9.9 | 315 | 22 | ABBA36403 | Human brain expres |
| C | 43 | 37.2 | 9.9 | 315 | 22 | AAKI17735 | Human bone marrow |
| C | 44 | 37.2 | 9.9 | 315 | 22 | AAKA43555 | Probe #14275 for g |
| C | 45 | 37.2 | 9.9 | 315 | 22 | AAI24342 | |

| | |
|----------|--|
| | ALIGNMENTS |
| | RESULT 1 |
| AAA28180 | 64..375 |
| ID | AAA28180 standard; cDNA: 375 BP. |
| XX | |
| AC | AAA28180; |
| XX | |
| XX | |
| DT | 29-JAN-2001 (first entry) |
| DE | |
| XX | platelet binding inhibitor protein Saratin encoding cDNA sequence. |
| XX | |
| KW | Saratin; medicinal leech; thromboembolic disease; intraocular lens |
| KW | collagen-dependent platelet adhesion inhibition; |
| KW | posterior capsule opacification; ss. |
| XX | |
| OS | Hirudo medicinalis. |
| XX | |
| FH | Key Location/Qualifiers |
| FT | CDS 64..375 |
| FT | /tag= a |
| FT | /product= "Saratin" |
| FT | /note= "platelet adhesion inhibitor protein" |
| FT | |

Same

xx Strittmatter W, Gnessow D, Hofmann U, Hemberger J, Fotev Z,
 pi Scheuble B;
 xx
 WP1: 2000-611629/58.
 DR P-PSDB: AAY94746.
 DR

PT New Saratin polypeptide and gene isolated from *Hirudo medicinalis* for blocking platelet adhesion, especially useful for treating or preventing thrombotic diseases, or for manufacturing a medicament for thromboembolic diseases -

Claim 4; Page 40; 46pp; English.

Saratin is a protein isolated from the saliva of the medicinal leech *Hirudo medicinalis*. Saratin is an inhibitor of collagen dependent platelet adhesion. The invention includes polynucleotide sequences encoding Saratin, an expression vector comprising the DNA sequence, a host cell transformed with the expression vector, antibodies immunospecific for Saratin, and methods for identifying Saratin agonists and for manufacturing a medicament for treating thromboembolic diseases. It is useful for preventing thrombotic diseases. Saratin is also useful for coating artificial surfaces, since use of Saratin renders them non-adhesive for cells and prevents the activation of cells. It may also be used for coating natural collagen surfaces. Furthermore, Saratin is useful for modifying intraocular lenses in order to lessen the thromboembogenicity of the lens material, for contacting the lens surface, or for covalent crosslinking to modify the lens material. The lens material is used for refractive anterior or posterior chamber ocular implants, which may be implanted into the eye. This new type of coating avoids problems contributed by stimulated cell growth. In combination with other medicaments that are for instance conferring cell death, Saratin coating helps to completely overcome posterior capsule opacification. The antibody immunospecific for Saratin, as well as Saratin itself, are useful for measuring samples derived from host cell cultures or from Saratin protein. The present sequence represents cDNA encoding the

Sequence 375 BP; 123 A; 62 C; 85 G; 105 T; 0 other;

| | | | | |
|---------------------------|---------|--------------------|--------|-------------|
| Query Match | 100.0%; | Score 375; | DB 21; | Length 375; |
| Best Local Similarity | 100.0%; | Pred. No. 3.3e-99; | | |
| Matches 375; Conservative | 0.0; | Mismatches | | |

0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTATTTCCTTGATTTCTCTTGGCTCGCAGCTTGCTGATCCACTACTCT 60
2 ATGAGTATTTCCTTGATTTCTCTTGGCTCGCAGCTTGCTGATCCACTACTCT 60

61 TCAGAGAAGCGTGAAGATTGTTGGACGTTTACGGAACAGAAATATACAGACTTCGAT 12

121 AATCTTTTAAGAGTCTCTGATCTTGACGAATCAAAAAACANGTTTCAAGACGGAG 18

181 TACTGCTACATCGTTTTTTTGAAGACACCCCGCAGGAG
121 AATCTTTTAAAGAAGTCCTTGATCTTTGACGAATTGCACAAAAAACAATCATGTTCACAGACGGAG

181 TACTGCTACATCGTTTTCAGACACCGTCAACAAGAAATGTTACTACAAATGTCGTTGAT 240

241 GGTGAAGACTTAGACCAGAAATTTTCTGTGCGAGAAACTTCACGGAATTTATTG 300

301 ACAGACTGCGAGGGTAAAGATGACAGTAAATCCCGGACCGGAAATTTATTTG 300
241 GGTGAAGAGTATTAGACCAAGAAATTTTGTGTGTCGACGAAAACTTACCGGAAATTTATTTG 300

301 ACAGACTGCAGGGTAAGATGCAGTATCGCCAGGTACAGGTGACGAGTCAGATGAA 360

361 GTTGATGAAGATTAA 375
|||||

RESULT 2
AAZ23891
ID AAZ23891 standard; DNA: 49999 BP.
XX

AC AAZ23891;

DT 25-JAN-2000 (first entry)
xx

Murine LOBO genomic DNA fragment 1.

XX spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

PN W09950284-A2.

| | |
|----|-------------|
| PD | 07-OCT-1999 |
| xy | |

20-MAR-1999; 99WO-EP02055

98DE-1013/99

(ROSE/) ROSENTHAL A.

Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

WPI; 1999-601320/51.

useful for treating and studying bone disorders

Example 3; Page 69-97; 391pp; German.

This invention describes novel nucleic acids (i.e. designated LOBO (long bones)) encoding proteins influencing bone development in mammals. The proteins of the invention reduce and/or inactivate bone extension (i.e. development), with exception of the skull and have osteoparathic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules, etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.

Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other,

| | | | | |
|---------------------------|--------|------------------|--------|---------------|
| Query Match | 12.2% | Score 45.8; | DB 20; | Length 49999; |
| Best Local Similarity | 49.8%; | | | |
| Matches 116: Conservative | | Pred. No. 0.011; | | |
| | | 0.011; | | |

| | conservative | U, mismatches | Indels | Gaps |
|---|--------------|---------------|--------|------|
| 143 ATCTTGACGAATGCAAAAAACATGCTTTCAGACGGAGTACTGCTACATCGCTTTTGAAG | | | | |
| 202 | | | | |

7333 ATTTGAGAGAAAAAATCTAATATTACAGATTCCAGACTTGTTATATAGAGACAGAGAG 7392

[illegible]

263 AATTGTGTGACGAAACTTCACGGAAATATTTTACAGACTCGAGGGTAAAGATG 322

7453 AAGAGAAGAGGAAAGAAGACAGACAGACAAGACAAGAACGAAGAAGAAAG 7512

323 CAGGTAATGCCGCAGGTACAGGTGACAGTCAAGTAGAAGTTGATGAAGATTAA 375
|| | | | | || | | | | || | | | |
7513 AAGACAAGAAGAAGAAGAAGAAGAACACACACACACACACACACACACACACACAC

RESULT 3
AAZ23896

Wed Feb 26 15:58:53 2003

us-09-936-737a-1.rng

AA23896 standard; DNA: 49999 BP.
 AA23896:
 25-JAN-2000 (first entry)
 Murine LOBO homologue genomic DNA fragment 2.
 LOBO: long bones; bone development; bone extension; skull; osteopathic;
 diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
 spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
 Mus musculus.
 MO950284-A2.
 07-OCT-1999.
 26-MAR-1999: 99NO-EP02055.
 27-MAR-1998: 98DE-1013799.
 (ROSE/) ROSENTHAL A.
 Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
 WPI: 1999-601320/51.
 Nucleic acids encoding proteins which influence bone development,
 useful for treating and studying bone disorders -
 Example 3; Page 161-189; 391pp; German.
 This invention describes novel nucleic acids (I; designated LOBO (long
 bones)) encoding proteins influencing bone development in mammals. The
 proteins of the invention reduce and/or inactivate bone extension (i.e.
 development), with exception of the skull and have osteopathic activity.
 The nucleic acid molecules, proteins and antibodies can be used in
 diagnosis or pharmaceutical compounds e.g. for gene therapy. The methods
 and nucleic acid molecules, etc. are useful for production of transgenic
 animals, especially a transgenic mouse for the study of diseases
 associated with bone development, e.g. spondyloepiphyseal dysplasia and
 achondroplasia. This sequence encodes the murine LOBO protein described
 in the method of the invention.
 Sequence 49999 BP: 13135 A; 11787 C; 10868 G; 14209 T; 0 other;
 Query Match 12.2%; Score 45.8; DB 20; Length 49999;
 Best Local Similarity 49.8%; Pred. No. 0.011;
 Matches 116; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
 Oy 143 ATCTTGACGAATGCAAAAACATGTTTCAAGCGAGTACTGCTACATGTTTGAAG 202
 Db 9318 ATTTTGAAGAAAAAATCTATATTACAGATTCACACTGTTATATAGAAAGAAAG 9377
 Oy 203 ACACGCTCAACAGAGATTTACTACATGCTGTGATGAGAGATTAGACCAAGAA 262
 Db 9378 AAG 9437
 Oy 263 AATTGTTGTCGAGCAAAATCTTCAAGAAATTTATTGACACAGTCGCGAGGTAAGATG 322
 Db 9438 AAG 9497
 Oy 323 CAGTAAATCGGCGAGTACAGTACAGTACAGTACAGTACAGTACAGTAAAGTTAA 375
 Db 9498 AAG 9550

RESULT 4
 ABL06114/C
 ID ABL06114 standard; cDNA: 15935 BP.
 AC ABL06114:
 AC

26-MAR-2002 (first entry)
 Drosophila melanogaster expressed polynucleotide SEQ ID NO 12824.
 Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical; gene; ss.
 Drosophila melanogaster.
 WO200171042-A2.
 27-SEP-2001.
 23-MAR-2001; 2001WO-US09231.
 23-MAR-2000; 2000US-191637P.
 11-JUL-2000; 2000US-0614150.
 (PEKE) PE CORP NY.
 Venter JC, Adams M, Li PWD, Myers EN;
 WPI: 2001-656860/75.
 P-PSDB: ABB62011.
 New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 Claim 1: SEQ ID NO 12824; 21bp + Sequence listing; English.
 The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 sequences (AB101840-AB116175) and the encoded proteins
 (AB157737-AB172072).
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 15935 BP: 4219 A; 3351 C; 3394 G; 4971 T; 0 other;
 Query Match 12.1%; Score 45.2; DB 23; Length 15935;
 Best Local Similarity 49.6%; Pred. No. 0.011;
 Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 Oy 141 TGATCTTGACGAATGCAAAAACATGTTTCAAGCGAGTACTGCTACATGTTTGA 200
 Db 10250 TAATGATGGGAGCGACGATGACGAGATGACAAAGAGAGAGAGAGAGATTAACGA 260
 Oy 201 AGACAGGCTCAACAGAGATTTACTACATGCTGTGATGAGAGATTAGACCAAGAA 10131
 Db 10190 AGAAGATATATATATATGACGAGAGATGACAGAGATGAGAGAGAGAGAGAG 10131
 Oy 261 AAAATTTGTTGTCGAGCAAAATCTTCAAGAAATTTATTGACACAGTCGCGAGGTAAGATG 10071
 Db 10130 CAATPATGATATAGAGAGAGAGATGACGAGAGATGACGAGAGATGATTAAGAGAGATTAACGA 10071
 Oy 321 TGCAGTAAATCGGCGAGTACAGTACAGTACAGTACAGTACAGTAAAGTTAA 374
 Db 10070 AGAAGATATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10017

RESULT 5
 AB054916
 ID AB054916 standard; cDNA: 2438 BP.
 AC AB054916:
 AC

Query Match

Length 2438;

| | |
|----------|------------------------|
| RESULT 6 | |
| AAS75452 | |
| ID | AAS75452 |
| XX | standard; cDNA; 372 bp |

XX (first entry)

Human: chromosome 11p15.5

Homologous recombination; medical imaging; 454 sequencing; gene therapy; forensic; diagnostics; genetic disorder; ssDNA

WO200175067-A2

2001WQ-US08631

43-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC

LIU C, Tang YT,

P-PSDB; ABG11265.

diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

NO 11256; 103pp; English

polypeptide (II) sequences. (I) is useful as hybridisation probes, and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags (ESTs). (II) is useful in gene therapy techniques to restore normal activity of (I) or to treat disease states involving a food supplement (II) and its binding partners are useful for imaging of sites expressing (II). (I) and (II) are useful in medical disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity.

| | |
|-----|--|
| | This cDNA sequence encodes a novel p160 (160.2) which is capable |
| CC | of activating transcription of a variety of genes upon activation of p62 |
| CC | and is capable of binding to the p62/p56lpx complex to modulate Ick |
| CC | function in a manner similar to p62. The genes transcribed in response to |
| CC | p160 activation likely include those of which are involved in T or B cell |
| CC | development/differentiation, T or B cell activation or production of T or |
| CC | B cell specific factors e.g., lymphokines or antibodies. This p160 |
| CC | polypeptide is also a substrate for serine/threonine kinase activity. |
| CC | p160 polypeptides can modulate degradation of cellular proteins e.g., cell |
| CC | cycle regulatory proteins stimulating expression of cell cycle dependent |
| CC | kinase inhibitors and arresting cell cycle progression at specific |
| CC | boundaries to thereby modulate cell proliferation. As p160 boosts B cell |
| CC | response it may be used to treat disorders where this is beneficial, e.g. |
| CC | infections by pathogenic microorganisms. p160 can be used to expand T |
| CC | cell populations for treating infectious diseases or cancer and p160 |
| CC | inhibitors could reduce B or T cell responses and may be used to treat a |
| CC | variety of autoimmune diseases, e.g. diabetes mellitus, arthritis, |
| CC | multiple sclerosis allergic reactions, Crohn's diseases etc. |
| SQ | Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T; 0 other; |
| OY | Query Match 10.7%; Score 40.2; DB 18; Length 3211; |
| | Best Local Similarity 52.0%; Pred. No. 0.17; 83; Indels 0; Gaps 0; |
| | Matches 90; Conservative 0; Mismatches 90; Matches 90; Gaps 0; |
| Db | 199 GAAGACACGCTCACAAAGCATGTACTTACATATGCCTTGATGTGAAGAATTGACCACA 258 |
| | |
| Oy | 2458 GAAGAAGAGAGCAAGCAAGAGAGAGAGAGCAGTTGGAGAAGAGAGAGATGAAGAG 2517 |
| | |
| Oy | 259 GAAAAAATTTGTTCGACCAAACCTTACCGAAAATTAATTGTCACAGCTCGAGGGGTAAA 318 |
| | |
| Ddb | 2518 GAATATTTTTGAAGAGAGCAAGAGAGAGAGAGAGAGTTTGAGAACAAATTGGAGAGAAA 2577 |
| | |
| Oy | 319 GATCCAGGTAAATGCCGCGATGACGTCAGTCAGTAGTACATGATGATGAAGA 371 |
| | |
| Ddb | 2578 GAAGGTGATTTAGAGAGAACAAAGAGAGATGACGAGGAGMAAGAA 2630 |
| | |
| | RESULT 8 |
| ID | AAT89345 standard; cdNA; 3901 BP. |
| AC | AAT89345: |
| XX | |
| DT | 11-MAR-1998 (first entry) |
| XX | |
| DE | Human p160 cdNA 160.1. |
| XX | |
| KW | p160; p62; cytoplasmic; T cell; B cell; development; activation; |
| KM | modulation; cellular response; cell proliferation; autoimmune disease; |
| KW | p56-Ick; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FH | Key Location/Oualifiers |
| FT | CDS 439..3846 |
| FT | /tag= a |
| FT | //product= p160 |
| FT | /note= "160.1" |
| NN | |
| PX | M09722255-A1. |
| PD | |
| Pf | 26-JUN-1997. |
| XX | |
| PF | 11-DEC-1996; 96MO-US19944. |
| XX | |
| PR | 19-DEC-1995; 95US-0574959. |
| PA | (DAND) DANA FARBER CANCER INST INC. |
| XX | |
| PI | Joung I, Shin J, Strominger JL, Vagstad RK; |
| XX | |
| DR | WPt. 1997-341351/31. |

DR P-PSDB; AAW31185.

XX CDNA encoding p62 and p160 and corresponding proteins - used in the
PT treatment of autoimmune disease and for T and B cell proliferation,
PT e.g. for treatment of tumours
XX
PS Claim 82; Fig 8; 175pp; English.

CC This cDNA sequence encodes a novel p160 (160.1) which is capable
CC of activating transcription of a variety of genes upon activation of p62
CC and is capable of binding to the p62/p56lck complex to modulate lck
CC function in a manner similar to p62. The genes transcribed in response to
CC p160 activation likely include those of which are involved in T or B cell
CC development/differentiation, T or B cell activation or production of T or
CC B cell specific factors e.g. lymphokines or antibodies. This p160
CC p160 polypeptide is also a substrate for serine/threonine kinase activity.
CC cycle regulatory proteins stimulating expression of cellular proteins e.g. cell
CC kinase inhibitors and arresting cell cycle progression at specific
CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
CC response it may be used to treat disorders where this is beneficial, e.g.
CC infections by pathogenic microorganisms. p160 can be used to expand T
CC cell populations for treating infectious diseases or cancer and p160
CC inhibitors could reduce B or T cell responses and may be used to treat a
CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis, etc.
CC multiple sclerosis allergic reactions, Crohn's diseases etc.

SQ Sequence 3901 BP; 781 A; 1183 C; 1128 G; 809 T; 0 other;

Query Match Best Local Similarity 10.7%; Score 40.2; DB 18; Length 3901;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAAGACAGGTCAACAGAGTGTACTACATGCTGTTGATGTAAGATTAGACCA 258
DB 3148 GAAGAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3207
QY 259 GAAATATTTGTTGCGAGCAAACTTCACGAAATTTATTTGACAGCTCGAGGTAA 318
DB 3208 GAAATATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3267
QY 319 GATCAGGTATGCGGAGGTGACAGTGCAGATGAATGATGATGA 371
DB 3268 GAAGGTGATTTAGAGAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAGAG 3320

RESULT 9

ABAA9279/C
ID ABAA9279 standard; DNA; 766 BP.

XX ABAA9279;

XX 01-FEB-2002 (first entry)

DE Human breast cell single exon nucleic acid probe #7974.

KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.

XX Homo sapiens.

OS WO200157271-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0633666.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WI; 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes

PS Claim 4; SEQ ID NO 7974; 327bp + sequence listing; English.

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;

Query Match Best Local Similarity 10.6%; Score 39.6; DB 22; Length 766;
Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 187 TACATCGTTTGAAGACAGCGTCAACAGAGTGTACTACATGCTGTTGATGTA 246
DB 687 TCCACACTTGAAGACAGCGTCAACAGAGTGTACTACATGCTGTTGATGTA 246
QY 247 GAGTTAGACCAAAATTTGTTGCGAGCAAACTTCACGAAATTTATTTGACAGC 306
DB 627 GATGAGAGGATGATTAATGATGATGATGATGATGATGATGATGATGATGATG 568
QY 307 TCGAGGTTAAAGATGCGGATGATGCGGAGGTACAGGTGATGATGATGATGATG 366
DB 567 GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 508
QY 367 GA 368
DB 507 GA 506

RESULT 10

ABA67193/C
ID ABA67193 standard; DNA; 766 BP.

XX ABA67193;

XX 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #15498.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157271-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 DR
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver -
 PS Claim 4; SEQ ID NO 15498; 639pp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14; Indels 0; Gaps 0;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTTGAAGACGCGTCAACAGGAATGTTACTACAAATGCTTGATGCTGAA 246
 Db TCCACACTTGAGAGCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628
 QY 247 GAGTTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTATTGACAGAC 306
 Db GATGCGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
 QY 307 TGGGAGGTTAAAGATGAGTATGCGGACGATGACAGTACAGTCAAGTGAAGTTGAT 366
 Db GACGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 11
 ABA34288/C
 ID ABA34288 standard; DNA; 766 BP.
 XX
 AC ABA34288;
 XX
 DE 23-JAN-2002 (first entry)
 XX
 XX Probe #12754 for gene expression analysis in human heart cell sample.
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 XX PN WO200157274-A2.
 XX

PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00666.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 DR
 PT Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts -
 PS Claim 4; SEQ ID NO 12754; 530pp; English.
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14; Indels 89; Gaps 0;
 Matches 93; Conservative 0; Mismatches 89; Indels 89; Gaps 0;
 QY 187 TACATCGTTTTGAAGACGCGTCAACAGGAATGTTACTACAAATGCTTGATGCTGAA 246
 Db TCCACACTTGAGAGCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628
 QY 247 GAGTTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTATTGACAGAC 306
 Db GATGCGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
 QY 307 TGGGAGGTTAAAGATGAGTATGCGGACGATGACAGTACAGTCAAGTGAAGTTGAT 366
 Db GACGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 12
 AA122114/C
 ID AA122114 standard; DNA; 766 BP.
 XX
 AC AA122114;
 XX
 DE 12-OCT-2001 (first entry)
 XX
 XX Probe #12047 for gene expression analysis in human cervical cell sample.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX

PD 09-AUG-2001.
 XX
 XX 29-JAN-2001; 2001WO-US00661.
 PF
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-476286/51.
 XX
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PS
 PS Claim 25; SEQ ID NO 7803; 322pp; English.
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
 SQ
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTTGAGACAGCGTCAACAGATGTTACTACATTCGTTGATGTTGAA 246
 DB 687 TCCACACTTGAAGCAACCATGAGATGATGATGATGATGATGATGATGATGATGATGAT 628
 QY 247 GAGTTAGACCAAGAAAATTTGTTGTCGACGAACCTTCACGGAATTAATTGACACAC 306
 DB 627 GATGAGAGAGATGATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 568
 QY 307 TCGGAGGTAAAGATGACGTAATGCGGACAGGTACAGGTGACAGAGTCAAGTGAAGTTGAT 366
 DB 567 GACGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 DB 507 GA 506
 XX
 XX RESULT 15
 XX ABM4123/c
 XX ID ABM4123 standard; DNA; 1944 BP.
 XX AC ABM4123;
 XX
 XX 01-FEB-2002 (first entry)
 XX
 XX Human breast cell single exon nucleic acid probe #2818.
 XX
 XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX

OS Homo sapiens.
 XX
 XX WO200157271-A2.
 PN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 30-JAN-2001; 2001WO-US00662.
 PF
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-496933/54.
 XX
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 PS
 PS Claim 1; SEQ ID NO 2818; 327pp + sequence listing; English.
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BR 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1944 BP; 466 A; 599 C; 349 G; 530 T; 0 other;
 SQ
 Query Match 10.6%; Score 39.6; DB 22; Length 1944;
 Best Local Similarity 51.1%; Pred. No. 0.21;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTTGAGACAGCGTCAACAGATGTTACTACATTCGTTGATGTTGAA 246
 DB 924 TCCACACTTGAAGCAACCATGAGATGATGATGATGATGATGATGATGATGATGATGAT 865
 QY 247 GAGTTAGACCAAGAAAATTTGTTGTCGACGAACCTTCACGGAATTAATTGACACAC 306
 DB 864 GATGAGAGAGATGATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 805
 QY 307 TCGGAGGTAAAGATGACGTAATGCGGACAGGTACAGGTGACAGAGTCAAGTGAAGTTGAT 366
 DB 804 GACGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 745
 QY 367 GA 368
 DB 744 GA 743
 XX

Search completed: February 25, 2003, 01:05:19
 Job time : 251 secs

•
•
•
•

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 00:54:10 ; Search time 46 Seconds
(without alignments)
2500.082 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagatattcttgatttc.....atgaagtgatgaagattaa 375

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--|
| C 1 | 42.2 | 11.3 | 7218 | 1 | US-08-232-463-14 Sequence 14, Appl |
| 2 | 40.2 | 10.7 | 3211 | 2 | US-08-574-959A-8 Sequence 8, Appl |
| 3 | 40.2 | 10.7 | 3211 | 4 | US-09-357-014-8 Sequence 8, Appl |
| 4 | 40.2 | 10.7 | 3901 | 2 | US-08-574-959A-6 Sequence 6, Appl |
| 5 | 40.2 | 10.7 | 3901 | 4 | US-09-357-014-6 Sequence 6, Appl |
| 6 | 39.6 | 10.6 | 3489 | 2 | US-08-728-323A-1 Sequence 1, Appl |
| 7 | 39.6 | 10.6 | 3489 | 4 | US-09-298-568-1 Sequence 1, Appl |
| 8 | 39.6 | 10.6 | 32207 | 2 | US-08-770-379-20 Sequence 20, Appl |
| C 9 | 39.6 | 10.6 | 32207 | 4 | US-08-757-669A-20 Sequence 20, Appl |
| C 10 | 39.6 | 10.6 | 32207 | 4 | US-09-230-371A-20 Sequence 20, Appl |
| 11 | 36 | 9.6 | 966 | 2 | US-08-766-738-2 Sequence 2, Appl |
| 12 | 36 | 9.6 | 966 | 4 | US-09-262-610-2 Sequence 2, Appl |
| 13 | 35 | 9.3 | 2694 | 2 | US-08-867-941-2 Sequence 2, Appl |
| 14 | 35 | 9.3 | 2694 | 4 | US-09-074-658-2 Sequence 2, Appl |
| 15 | 35 | 9.3 | 7650 | 2 | US-08-867-941-1 Sequence 1, Appl |
| 16 | 35 | 9.3 | 7650 | 4 | US-09-074-658-1 Sequence 1, Appl |
| 17 | 35 | 9.3 | 8920 | 2 | US-08-446-855A-1 Sequence 1, Appl |
| 18 | 35 | 9.3 | 8920 | 4 | US-09-150-741-1 Sequence 1, Appl |
| 19 | 34.8 | 9.3 | 5361 | 4 | US-08-973-462-2 Sequence 1, Appl |
| 20 | 34.8 | 9.3 | 6152 | 4 | US-08-973-462-1 Sequence 1, Appl |
| 21 | 34.6 | 9.2 | 2518 | 4 | US-09-433-699-3 Sequence 3, Appl |
| 22 | 34.2 | 9.1 | 5852 | 1 | US-07-867-106-2 Sequence 2, Appl |
| 23 | 34 | 9.1 | 759 | 1 | US-08-466-603-4 Sequence 4, Appl |
| 24 | 34 | 9.1 | 759 | 1 | US-08-314-503A-4 Sequence 4, Appl |
| 25 | 34 | 9.1 | 759 | 1 | US-08-468-066-4 Sequence 4, Appl |
| 26 | 34 | 9.1 | 759 | 1 | US-08-466-717-4 Sequence 4, Appl |
| 27 | 34 | 9.1 | 759 | 3 | US-08-466-743-4 Sequence 4, Appl |

| | | | | | |
|----|------|-----|------|---|---|
| 28 | 34 | 9.1 | 759 | 5 | PCT-US95-12414-4 Sequence 4, Appl |
| 29 | 34 | 9.1 | 980 | 1 | US-08-466-603-3 Sequence 3, Appl |
| 30 | 34 | 9.1 | 980 | 1 | US-08-314-503A-3 Sequence 3, Appl |
| 31 | 34 | 9.1 | 980 | 1 | US-08-468-066-3 Sequence 3, Appl |
| 32 | 34 | 9.1 | 980 | 2 | US-08-466-717-3 Sequence 3, Appl |
| 33 | 34 | 9.1 | 980 | 2 | US-08-466-743-3 Sequence 3, Appl |
| 34 | 34 | 9.1 | 980 | 4 | PCT-US95-12414-3 Sequence 3, Appl |
| 35 | 33.8 | 9.0 | 4518 | 4 | US-08-961-527-121 Sequence 121, App |
| 36 | 33.6 | 9.0 | 2295 | 1 | US-08-375-300-3 Sequence 3, Appl |
| 37 | 33.6 | 9.0 | 2295 | 1 | US-09-177-431-3 Sequence 3, Appl |
| 38 | 33.6 | 9.0 | 2295 | 5 | PCT-US95-16930-3 Sequence 3, Appl |
| 39 | 33.6 | 9.0 | 4080 | 1 | US-08-375-300-1 Sequence 1, Appl |
| 40 | 33.6 | 9.0 | 4080 | 3 | US-09-177-431-1 Sequence 1, Appl |
| 41 | 33.6 | 9.0 | 4080 | 5 | PCT-US95-16930-1 Sequence 1, Appl |
| 42 | 33.4 | 8.9 | 1137 | 4 | US-09-134-001C-657 Sequence 657, App |
| 43 | 33.4 | 8.9 | 1727 | 4 | US-09-071-035-295 Sequence 295, App |
| 44 | 33.4 | 8.9 | 1839 | 4 | US-09-071-035-293 Sequence 293, App |
| 45 | 33.2 | 8.9 | 4599 | 1 | US-08-431-080-27 Sequence 27, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHTEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 11.3%; Score 42.2; DB 1; Length 7218;


```

: GENERAL INFORMATION:
: APPLICANT: Jaekyoon Shin, Inslil Joung, Ratna K. Vadlamudi
: and Jack L. Strominger
: TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
: AND USES THEREFOR
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/357,014
: FILING DATE: 19-Jul-1999
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/574,959
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Mandragouras, Amy E.
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: DFN-008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3901 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 439..3847
: ? ? ? SEQUENCE DESCRIPTION: SEQ ID NO: 6:
: US-09-357-014-6
:
: Query Match 10.7%; Score 40.2; DB 4; Length 3901;
: Best Local Similarity 52.0%; Pred. No. 0.0073;
: Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
:
: QY 199 GAAGACACGCGTCAACGAGAAATGTTACTACATGTCGTTGATGGTGAAGAGTAGACCAA 258
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 3148 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3207
:
: QY 259 GAAAAATTGTTGTCGACGAAAACTTCACGGAATAATTATTTGACAGACTGCGAGGGGTAA 318
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 3208 GAATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3267
:
: QY 319 GATGCGAGGTAAATGCGGACGAGTACAGGTGACGAGTCACATGAAGTTGATGAAGA 371
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 3268 GAAGGTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3320
:
: RESULT 6
: US-08-728-323A-1
: Sequence 1, Application US/08728323A
: Patent No. 5948676
:
: GENERAL INFORMATION:
: APPLICANT: Chany, Yuan
: APPLICANT: Bohenzky, Roy A.
: APPLICANT: Russo, James J.
: APPLICANT: Edelman, Isidore S.
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: Immediate Early Protein From Kaposi's
: Sarcoma-Associated Herpesvirus, DNA

```

```

; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MS/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1

Query Match          10.6%; Score 39.6; DB 2; Length 3489;
Best Local Similarity 47.2%; Pred. No. 0.011;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

; 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAATGCAAAAAACATGTTTCAAGAGC 177
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 871 GAAGAACATCCAAAGATGAAAAAGATGTCGGAATAATATAGCTGGCCGAGATTAAT 930
;
; QY 178 GAGTACTGCTACATCGTTTGAAGACACGGTCAACAGAACTTCTACAAATGTCGT 237
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 931 GGGGACACGAGATTCAGCAAGAAATGTCGGAATAATATAGCTGGCCGAGATTAAT 930
;
; QY 238 GATGCTGAAGAGTTAGACCAAGAAATTTGTTGTCACGAAACTTCACGAAATTAAT 297
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 991 GATGAGGAGGACAGAGACAGATGACGAGAGAGATGACGAGAGAGATGACGAGAG 1050
;
; QY 298 TTGACAGACTCGAGGCTGAAGATGCGAGTAAATCGCGCAGTACAGTACAGTACAGAT 357
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1051 GAGGATGACGAGGAGATGACGAGAGAGATGACGAGAGAGATGACGAGAGATGACGAG 1110
;
; QY 358 GAAGTTGATGAAGA 371
;   ||| ||| ||| |||
; Db 1111 GAGGATGACGAGGA 1124

RESULT 7
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestras, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R

```

```

; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-09-298-568-1

Query Match          10.6%; Score 39.6; DB 4; Length 3489;
Best Local Similarity 47.2%; Pred. No. 0.011;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

; 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAATGCAAAAAACATGTTTCAAGAGC 177
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 871 GAAGAACATCCAAAGATGAAAAAGATGTCGGAATAATATAGCTGGCCGAGATTAAT 930
;
; QY 178 GAGTACTGCTACATCGTTTGAAGACACGGTCAACAGAACTTCTACAAATGTCGT 237
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 931 GGGGACACGAGATTCAGCAAGAAATGTCGGAATAATATAGCTGGCCGAGATTAAT 930
;
; QY 238 GATGCTGAAGAGTTAGACCAAGAAATTTGTTGTCACGAAACTTCACGAAATTAAT 297
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 991 GATGAGGAGGACAGAGACAGATGACGAGAGAGATGACGAGAGAGATGACGAGAG 1050
;
; QY 298 TTGACAGACTCGAGGCTGAAGATGCGAGTAAATCGCGCAGTACAGTACAGTACAGAT 357
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1051 GAGGATGACGAGGAGATGACGAGAGAGATGACGAGAGAGATGACGAGAGATGACGAG 1110
;
; QY 358 GAAGTTGATGAAGA 371
;   ||| ||| ||| |||
; Db 1111 GAGGATGACGAGGA 1124

RESULT 8
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525

```

| | | | |
|----|-------|--|-------|
| OY | 118 | GAAATCTTTAAGGAAGTCCTCGATTCTTGACGAAATGCATAAAAAACAATGTTCMAAGC | 177 |
| Dd | 21126 | GAGAACCATCCAACGAATGAAAAGCATTCTCCGAAATTAATCAAGGCTGGCGAGATTAAT | 21067 |
| OY | 178 | GACTACTCGTACATCGTTTTTGAAGCACGSCCAACAAGSANTGTACTACATGTCTGT | 237 |
| Dd | 21066 | GGGACACACAGATTTCGACAGGAAMCTCAGSTTGCACAGATGACATTCGACATAAGGAT | 21007 |
| OY | 238 | GATGCTGAAGGTTTAGACCACAGAAAAATTTGTTCTCGASNAATCTCCAGGAANAATTAT | 297 |
| Dd | 21006 | GATGAGGAGGACGACGAGACACAGTGAAGGAGGACGAGAGAGGATGACGAGAGGATGACGAG | 20947 |
| OY | 298 | TTCGACACTGCCAGGCTAAACATGCAGGTAATCCGCGCAGGTACAGGTGACGAGTCAGAT | 357 |

Dd 20946 GAGGATACGAGCAGGATTACCGAGGATGACGAGGATGACGAGGGAGTGAACGAG 20873
 ||| |||| | | | | | | | | | | | | |
Oy 358 GAAGTTGATGAGA 371
 || | ||| | | |
Dd 20886 GAGGATGCACGAGA 20873

RESULT 11
UC-09-766

```

US-08-766-738-2
: Sequence 2, Application US/08766738
: Patent No. 5916749
:
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Goli, Suvya K.
: TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
:
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/766,738
: FILING DATE: Herewith
: CLASSIFICATION: 530
:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER:
:
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0177 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 966 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: Consensus
:
: CLONE: 1013361
:
US-08-766-738-2

```

| | | | | |
|-----------------------|--------|------------------|-------|----------------------------------|
| Query Match | 9.6%; | Score 36; | DB 2; | Length 966; |
| Best Local Similarity | 49.5%; | Pred. No. 0.077; | | |
| Matches | 90; | Conservative | 0; | Mismatches 92; Indels 0; Gaps 0. |

[illegible]

Db 763 GA 764

RESULT 12
US-09-262

```

US-09-262-610 2
Sequence 2, Application US/09262610
Patent No. 6428949
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goli, Suvya K.
TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,610
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/766,738
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0177 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: 1813361
US-09-262-610-2

```

| | | | | |
|-----------------------|--------|------------------|--------|----------------|
| Query Match | 9.68; | Score 36; | DB 4; | Length 966; |
| Best Local Similarity | 49.58; | Pred. No. 0.077; | | |
| Matches | 90; | Conservative | 0; | Mismatches 92; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

| QY | 193 | GTTTGTGAAGACACGCTCAACAGCAGTGTACTACAACTGTGGTATGCTGTAAGAGTTA | 253 |
|----|-----|---|-----|
| Db | 583 | GTTGATTAAAGAGAGAGAGACGGAGAAGAGAAATGAGAGAACGACGACATGAGGAT | 642 |
| QY | 253 | GACCACAGAAAAATTTGTTGTCGACGAAAACCTTCACGGAATTTATTTCACGACTCGCAG | 312 |
| Db | 643 | GGTGAGAGAGAGAGAGTTTGATGTAGAAAGATGATGAAATATGTAAGAGGGATGAG | 702 |
| QY | 313 | GCTAAGATGCGAGGTATATGCGGACAGTGCAGGTGACGAGTCAGATGAAATGTTGATGACAT | 372 |
| Db | 703 | GACGACCATGAACTCAGTGTAGAGAGAAAGCAATTTCGACTTGTATGTAAGAAAGATGAAGAT | 762 |
| QY | 373 | TA 374 | |
| Db | 763 | GA 764 | |

Wed Feb 26 15:58:54 2003

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jdb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-1

```

```

Query Match          9.3%; Score 35; DB 2; Length 7650;
Best Local Similarity 52.2%; Pred. NO. 0.42;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCGTCACAGAGATGTTACTACAT 231
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2157 AAAAAGACAAAGGTTATAGCAATATGAGAAACCATCAGAAAAAGGCCATCAGAT 2216
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 GTGCTGATGAGTGAAGAGTGAACCAAGAAAATTGTGTGACGAAACTTCACGGA 291
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2217 TATCTGTAAACCGAGACTTCACCCAGAA-----GATGATGACGATGTTGACCGCA 2270
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 292 AATTATTGACAGACTGGAGGGTAAAGATCCAGTAAATGCGGACGATACAGGTGACGAG 351
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2271 TCTGATGATTCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2330
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 TCAGATGAGTGTGATGATGAT 372
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2331 TCACAGATGATGACGAGAT 2351
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: February 25, 2003, 01:50:29
 Job time : 71 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 01:05:25 ; Search time 61 seconds

(without alignments)
3452.701 Million cell updates/sec

Title: US-09-936-737a-1

Perfect score: 375
Sequence: 1 atgaagtattcttgatttc.....atgaagtgtgaagatttaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 segs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|---------------|-------|----------------|--------|----|---------------------|
| 1 | 39.6 | 10.6 | 766 | 10 | US-09-864-761-19608 |
| 2 | 39.6 | 10.6 | 1944 | 10 | US-09-864-761-2825 |
| 3 | 37.2 | 9.9 | 315 | 10 | US-09-864-761-2123 |
| 4 | 36.6 | 9.8 | 499 | 10 | US-09-783-590-5790 |
| 5 | 36 | 9.6 | 381 | 9 | US-10-015-219-700 |
| 6 | 36 | 9.6 | 381 | 10 | US-09-777-564-700 |
| 7 | 36 | 9.6 | 966 | 9 | US-10-213-700-2 |
| 8 | 35.8 | 9.5 | 204 | 10 | US-09-864-761-21008 |
| 9 | 35.8 | 9.5 | 474 | 10 | US-09-864-761-4255 |
| 10 | 35.4 | 9.4 | 876 | 10 | US-09-770-445-544 |
| 11 | 35.4 | 9.4 | 1092 | 9 | US-09-938-842A-1125 |
| 12 | 35 | 9.3 | 267 | 10 | US-09-878-574-15076 |
| 13 | 35 | 9.3 | 563 | 10 | US-09-864-761-13293 |
| 14 | 34.8 | 9.3 | 327 | 10 | US-09-864-761-28059 |
| 15 | 34.8 | 9.3 | 1626 | 10 | US-09-764-848-19 |
| 16 | 34.8 | 9.3 | 5361 | 9 | US-09-742-096-2 |
| 17 | 34.8 | 9.3 | 6152 | 9 | US-09-742-096-1 |
| 18 | 34.6 | 9.2 | 438 | 10 | US-09-864-761-4988 |
| 19 | 34.6 | 9.2 | 542 | 10 | US-09-864-761-12239 |

| | | | | | | |
|----|------|-----|--------|----|---------------------|--------------------|
| 20 | 34.6 | 9.2 | 8391 | 9 | US-10-151-736-3 | Sequence 3, Appl1 |
| 21 | 34.6 | 9.2 | 8493 | 9 | US-10-151-736-5 | Sequence 5, Appl1 |
| 22 | 34.4 | 9.2 | 846 | 10 | US-09-815-242-8918 | Sequence 8918, Ap |
| 23 | 34.4 | 9.2 | 1282 | 9 | US-10-002-344A-89 | Sequence 89, Appl1 |
| 24 | 34.2 | 9.1 | 381 | 10 | US-09-864-761-21610 | Sequence 21610, A |
| 25 | 34.2 | 9.1 | 659158 | 9 | US-09-771-208-20 | Sequence 20, Appl1 |
| 26 | 33.8 | 9.0 | 1824 | 10 | US-09-815-242-9281 | Sequence 9281, Ap |
| 27 | 33.8 | 9.0 | 1824 | 10 | US-09-815-242-9584 | Sequence 9584, Ap |
| 28 | 33.6 | 9.0 | 305 | 10 | US-09-864-761-19262 | Sequence 19262, A |
| 29 | 33.6 | 9.0 | 496 | 10 | US-09-864-761-2534 | Sequence 2534, Ap |
| 30 | 33.4 | 8.9 | 420 | 10 | US-09-864-761-23266 | Sequence 23266, A |
| 31 | 33.4 | 8.9 | 451 | 10 | US-09-864-761-6554 | Sequence 6554, Ap |
| 32 | 33.4 | 8.9 | 451 | 10 | US-09-864-761-14152 | Sequence 14152, A |
| 33 | 33.4 | 8.9 | 7657 | 10 | US-09-070-927A-33 | Sequence 33, Appl1 |
| 34 | 33.2 | 8.9 | 157 | 10 | US-09-864-761-29856 | Sequence 29856, A |
| 35 | 33.2 | 8.9 | 423 | 9 | US-09-938-842A-592 | Sequence 592, App |
| 36 | 33.2 | 8.9 | 58985 | 9 | US-09-901-152-3 | Sequence 3, Appl1 |
| 37 | 33 | 8.8 | 1959 | 10 | US-09-864-761-4012 | Sequence 4012, Ap |
| 38 | 33 | 8.8 | 2920 | 10 | US-09-801-574-87 | Sequence 87, Appl1 |
| 39 | 33 | 8.8 | 5711 | 9 | US-09-734-672-1 | Sequence 1, Appl1 |
| 40 | 33 | 8.8 | 5711 | 9 | US-09-734-672-3 | Sequence 3, Appl1 |
| 41 | 33 | 8.8 | 5711 | 9 | US-09-734-672-5 | Sequence 5, Appl1 |
| 42 | 33 | 8.8 | 5711 | 9 | US-09-982-828-1 | Sequence 1, Appl1 |
| 43 | 33 | 8.8 | 5711 | 9 | US-09-982-828-3 | Sequence 3, Appl1 |
| 44 | 33 | 8.8 | 5711 | 9 | US-09-982-828-5 | Sequence 5, Appl1 |
| 45 | 33 | 8.8 | 5711 | 9 | US-10-022-819-1 | Sequence 1, Appl1 |

ALIGNMENTS

RESULT 1
US-09-864-761-19608/c
Sequence 19608, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

| | | | | |
|--------------------------|--------|------------------|-----------|-------------|
| Query Match | 10.6%; | Score 39.6; | DB 10; | Length 766; |
| Best Local Similarity | 51.1%; | Pred. No. 0.051; | | |
| Matches 93; Conservative | 0; | Mismatches 89; | Indels 0; | Gaps 0 |

| | | | |
|----|-----|---|-----|
| QY | 187 | TACATGTTTTTGAAGACACCGGTCAACACAGAAATGTACTACATCTGCTATGTTGCA | 246 |
| Db | 687 | TCACACCTTGAGGACACACCATGAGAGATGATGTGTGATGACAGATGACATGATGCTGT | 622 |
| QY | 247 | GAGTTAGACCAAGAAAAATTTGTTGTCGACGAAAACCTTCACGGAATTAATTTCACAGAC | 308 |
| Db | 627 | GATGACAGAGATGATATAATATGATGACGAGATGATGATTAATGATATGTTGCGGAGAG | 566 |
| QY | 307 | TGCGAGAGTTAAAGATGCACGTTAATGCCGACAGGTCACAGGTGACGAGTCACAGTAAAGTTGAT | 366 |
| Db | 567 | GACGACGAGCAAAATGATGCTGATGAGTGCATGATGAAAAAATGATATATGACGAGAAATCAGAT | 508 |
| QY | 367 | GA 368 | |
| Db | 507 | GA 506 | |

```

RESULT 2
US-09-864-761-2825/C
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Weisheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30

```

```

1 PRIOR APPLICATION NUMBER: PCT/US01/00667
2 PRIOR FILING DATE: 2001-01-30
3 PRIOR APPLICATION NUMBER: PCT/US01/00664
4 PRIOR FILING DATE: 2001-01-30
5 PRIOR APPLICATION NUMBER: PCT/US01/00669
6 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00665
8 PRIOR FILING DATE: 2001-01-30
9 PRIOR APPLICATION NUMBER: PCT/US01/00668
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00663
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00662
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00661
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00670
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: US 60/234,687
20 PRIOR FILING DATE: 2000-09-21
21 PRIOR APPLICATION NUMBER: US 09/608,408
22 PRIOR FILING DATE: 2000-06-30
23 PRIOR APPLICATION NUMBER: US 09/774,203
24 PRIOR FILING DATE: 2001-01-29
25 NUMBER OF SEQ ID NOS: 49117
26 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1.
27 SEQ ID NO 2825
28 LENGTH: 1944
29
30
31

```

```

: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL008720.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
: OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
: OS-09-864-761-2825

```

| | | | | |
|-----------------------|-----------------|------------------|-----------|--------------|
| Query Match | 10.6% | Score 39.6; | DB 10; | Length 1944; |
| Best Local Similarity | 51.1%; | Pred. No. 0.075; | | |
| Matches 93; | Conservative 0; | Mismatches 89; | Indels 0; | Gaps 0 |

| | | | |
|----|-----|---|-----|
| QY | 187 | TACATCGTTTTGAAGAACGCTCACACAGAAATTGTACTACAACTGCTTAATGGTGA | 24 |
| Dd | 924 | TCACACCTTGAGGACAACCATGAGATGATGTGTGATGATGACATGAAGATGTTGCT | 865 |
| QY | 247 | GAGTTAGACCAAGAAAAATTTTGTGTGCAGCAAAAACCTTCACGAAATAATTTCACAGC | 306 |
| Dd | 864 | GATGAGAGAGATGATTAATTAATGATGACGATGATGATGATTAATGATATGATGGAGAG | 808 |
| QY | 307 | TGCGAGGGTAAAGATGCAGTGAATGCCGACAGTACAGGTACAGTCCACATGAAGTTGAT | 366 |
| Dd | 804 | GACGAGAGAANAATGATGTGATGATGATGATGATGAATAAATGATTAATGAGAGAAATGAGAT | 745 |
| QY | 367 | GA 368 | |
| Dd | 744 | GA 743 | |

```

RESULT 3
US-09-864-761-21723
; Sequence 21723; Application US/09864761
; Patent No. US20020048763A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
;
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;
; FILE REFERENCE: Aeomica-X-1

```

```

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ. ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ. ID NO 21723
LENGTH: 315
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL035353.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: U28921.1, EVAL0E 6.50e-01
US-09-864-761-21723

```

| | Query Match | Similarity | Score | DB | Length |
|------------|-------------|--------------|-------|------------|--------|
| Best Local | 9.9% | 50.0% | 37.02 | 10 | 315 |
| Matches | 93 | Conservative | 0 | Mismatches | 93 |
| | | | | Indels | 0 |
| | | | | Gaps | 0 |

```

Db      172  GAACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 231
Oy      370  GATTAA 375
         || ||
Db      232  GAAGAA 237

RESULT 4
US-09-783-590-5790
; Sequence 5790, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2.
; FILE REFERENCE: PD-16.2c1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIORITY FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5790
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (127)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (211)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (233)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (262)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (330)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (337)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (345)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (346)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (348)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)

```

```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (417)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (423)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (439)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (456)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (467)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (468)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (469)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (470)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (476)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (492)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (497)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5790
Query Match          9.8%; Score 36.6; DB 10; Length 499;

```

```

Best Local Similarity 50.6%; Pred. No. 0.31;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 31 CTCGACGCTTGCATCTCACTACTCTTCTTCAGAAAGAGTGAAGTTGTGACGCTT 90
    || || || || || || || || || || || || || || || || || || ||
Db 10 CTCAGATTCFAGTAATCTAAACCCCTGCCAGGTGACCAAGATCAAGTTGTGTAC 69

QY 91 TACCGAAGCAAGAAATATACAGACTTGCATTAATCTTTTAAAGTCTGTATCTTGAC 150
    || || || || || || || || || || || || || || || || || || ||
Db 70 TCCAGACACAGAAAGAGATGCTTAGATAGACAGTAAGAAATGTATATCAATAT 129

QY 151 GAATGCAAAAAACATGTTTCAAGACGAGTACTGCTATCGTTTGTAG 202
    || || || || || || || || || || || || || || || || || || ||
Db 130 CATTTCATGTAATTAATCTTTTAATTTGAAACATTAGCAATGCTGTAGATG 181

RESULT 5
US-10-015-219-700
; Sequence 700, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493c1
; CURRENT APPLICATION NUMBER: US/10/015,219
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 66, 201, 213, 225, 251, 255, 261, 262, 265, 280, 290, 303,
; LOCATION: 313, 322, 339, 373
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-700

Query Match          9.6%; Score 36; DB 9; Length 381;
Best Local Similarity 49.5%; Pred. No. 0.42;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAGACACGCTCAACAAGAAATGTACTACATGCTGTGATGTAAGATT 252
    || || || || || || || || || || || || || || || || || || ||
Db 21 GTGGATGAGAGGAGGAGGACGAGAGGAGGAAGATGAGAGACNAGACGATGAT 80

QY 253 GACCAAGAAAAATTTGTGTCGACGAAACTTCACGAAATTTATTTGACAGCTGCGAG 312
    || || || || || || || || || || || || || || || || || || ||
Db 81 GGTGAAGAGAGAGAGTTTATGATAGAGATGATGAAGATGAAGATGTAGAACGGATGAG 140

QY 313 GGTAAAGATGACGATGATGCGCAGGTGACGATGACGATGATGAAGTTGATGATGAT 372
    || || || || || || || || || || || || || || || || || || ||
Db 141 GAGCAGCATGAGTACGATGAGGAGGAAAGAAATTTGACCTGTGATGAGAAAGATGAT 200

QY 373 TA 374
Db 201 NA 202

RESULT 6
US-09-777-564-700
; Sequence 700, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05

```

NUMBER OF SEQ ID NOS: 1730
 SOFTWARE: FastSeq for Window Version 4.0
 SEQ ID NO 700
 LENGTH: 381
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc-feature
 LOCATION: (1)...(381)
 OTHER INFORMATION: n = A,T,C or G
 US-09-777-564-700

Query Match
 Best Local Similarity 49.5%; Pred. No. 0.42; Length 381;
 Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTGAAGACACGGTCAACAGCAATGTTACTCAATGTCGTGATGTTGAAGAGTTA 252
 DB 21 GGGATGAAGAT 80
 QY 253 GACCAAGAAAAATTTGTCGACGAAACTTCAACGAAATTTATTTGACAGACTGCGAG 312
 DB 81 GGTGAAGAAAGAT 140
 QY 313 GGTAAAGATCGAGTAAATGCGGAGGTACAGAGTCAAGTGAAGTGAAGTGAAGAT 372
 DB 141 GACGACGATGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 200
 QY 373 TA 374
 DB 201 NA 202

RESULT 7

US-10-213-700-2
 Sequence 2, Application US/10213700
 Publication No. US20030022332A1
 GENERAL INFORMATION:

APPLICANT: Bandman, Olga
 TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/213.700
 FILING DATE: 06-Aug-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/766,738
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0177 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166

TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 966 base pairs
 TYPE: nucleic acid

STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: Consensus
 CLONE: 1813361
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-213-700-2

Query Match
 Best Local Similarity 49.5%; Pred. No. 0.61; Length 966;
 Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTGAAGACACGGTCAACAGCAATGTTACTCAATGTCGTGATGTTGAAGAGTTA 252
 DB 583 GTTGATNAAGANGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 642
 QY 253 GACCAAGAAAAATTTGTCGACGAAACTTCAACGAAATTTATTTGACAGACTGCGAG 312
 DB 643 GGTGAAGAAAGAT 702
 QY 313 GGTAAAGATCGAGTAAATGCGGAGGTACAGAGTCAAGTGAAGTGAAGTGAAGAT 372
 DB 703 GACGACGATGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 762
 QY 373 TA 374
 DB 763 GA 764

RESULT 8

US-09-864-761-21008/C
 Sequence 21008, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aecm1ca-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30

```

PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21008
LENGTH: 204
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121580.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97
OTHER INFORMATION: NT HIT: AP155827.1, EVALUATE 1.60e+00
OTHER INFORMATION: EST_HUMAN HIT: BE889896.1, EVALUATE 2.80e+00
US-09-864-761-21008

```

```

Query Match          9.5%; Score 35.8; DB 10; Length 204;
Best Local Similarity 54.1%; Pred. No. 0.37;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

QY 238 GATGCTGAAGAGTGAACCAAGAAATTTGTGTGTCACGAAACTTCACGGAATTTAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 GATGATGAAGGTGATGAATGAAGATGAAGAAAGAAAGATGTTAAGATGAA 112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 TTGACAGACTGCGAGGTTAAAGATGATGCGGCGAGTACAGGTGACGATCAGAT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 GATGAAGAAGATGAATAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGAA 52
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 GAAGTTGATGAAGAT 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 GATGAATAATGAAGAT 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
US-09-864-761-4255/C
Sequence 4255, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US/09/864,761
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30

```

```

PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4255
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121580.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97
US-09-864-761-4255

```

```

Query Match          9.5%; Score 35.8; DB 10; Length 474;
Best Local Similarity 54.1%; Pred. No. 0.52;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

QY 238 GATGCTGAAGAGTGAACCAAGAAATTTGTGTGTCACGAAACTTCACGGAATTTAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 GATGATGAAGGTGATGAATGAAGATGAAGATGAAGAAAGAAAGATGTTAAGATGAA 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 TTGACAGACTGCGAGGTTAAAGATGATGCGGCGAGTACAGGTGACGATCAGAT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 GATGAAGAAGATGAATAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 GAAGTTGATGAAGAT 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 GATGAATAATGAAGAT 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
US-09-770-445-544/C
Sequence 544, Application US/09770445
Patent No. US20020023281A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.

```

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 52 |
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|

| | | | | |
|-----------------------|--------------|----------------|---------------|-------------|
| Query Match | 9.4% | Score 35.4 | DB 9 | Length 1092 |
| Best Local Similarity | 52.3% | Pred. No. 0.96 | | |
| Matches 78 | Conservative | 0 | Mismatches 71 | Indels 0 |
| | | | | Gaps 0 |

```

? RESULT 12
? US-09-878-574-15076
? Sequence 15076, Application US/09878574
? Patent No. US20020110548A1
? GENERAL INFORMATION:
? APPLICANT: Byrum, Joseph R.
? APPLICANT: La Rosa, Thomas J.
? APPLICANT: Thompson, Michael D.
? TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
? TITLE OF INVENTION: Plants
? FILE REFERENCE: 38-21(15401)B
? CURRENT APPLICATION NUMBER: US/09/878,574
? CURRENT FILING DATE: 2001-12-21
? PRIOR APPLICATION NUMBER: 09/333,535
? PRIOR FILING DATE: 1999-06-14
? NUMBER OF SEQ ID NOS: 15775
? SEQ ID NO 15076
? LENGTH: 267
? TYPE: DNA
? ORGANISM: Glycine max
? OTHER INFORMATION: Clone ID: 701069621H1
? US-09-878-574-15076

```

```

; RESULT 13
; US-09-864-761-13293
; Sequence 13293, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

```

[illegible]

RESULT 15
US-09-764-848-19
; Sequence 19, Application US/09764848
; Patent No. US20020077270A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT208
; CURRENT APPLICATION NUMBER: US/09/764,848
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-848-19

Query Match

Best Local Similarity 9.3%; Score 34.8; DB 10; Length 1626;
Matches 78; Conservative 1; Mismatches 73; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 217 | GAATGTTACTACAAATGTCGTGATGCTGAAGATTAGACCAAGAAATTTGTTGTCGAC | 276 |
| | | | |
| DB | 459 | GCATCTATATTAAGGAAGTTGAAGATTAAGATCAAGAGGAGAGAGAGATGAGAT | 518 |
| | | | |
| QY | 277 | GAATCTCAGCGAAATTTATTGACAGACTGCGAGSGGTAAAGATGCAGTAAATGGGCA | 336 |
| | | | |
| DB | 519 | GAAGATCTTTCTAAATTAATAGTAAAGATGAGGATGAGGATGAAGATGACGCTGATCTCTCA | 578 |
| | | | |
| QY | 337 | GGTACAGGTGAGAGCTCAGATGAGTGAATGATGA | 368 |
| | | | |
| DB | 579 | AAATTAATCTTGATGCCAGTGAAGAGAGAGA | 610 |
| | | | |

Search completed: February 25, 2003, 02:24:22
Job time : 65 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:23:10 : Search time 22 Seconds
(without alignments)
3277.310 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 674
Sequence: 1 atgaagtattctgtattc.....atgaagtgtatgaatataa 375

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+_n2p.model -DEV=xlh
-Q=/cgn2.1/USPTO/US093673/1/runat_14022003.100558.1407/app-query.fasta.1.519
-DB=PIR_73 -OFMT=fastan -SUFFIX=pr -MINMATCH=0.1 -LOOPEXT=0
-UNIT8-bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -ALIGN=15 -MODE=LOCAL
-ORFMT=pr -NCM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US093673/1/cgn2.1.16.ernal_14022003.100558.1407 -NCP=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 88.5 | 13.1 | 147 | 2 A42435 | leech antiplatelet |
| 2 | 84.5 | 12.5 | 164 | 2 T33824 | hypothetical prote |
| 3 | 82.5 | 12.2 | 412 | 2 A55320 | immunophilin FKBP4 |
| 4 | 82 | 12.6 | 637 | 2 T00548 | hypothetical prote |
| 5 | 82 | 12.2 | 907 | 2 T27317 | hypothetical prote |
| 6 | 80.5 | 11.9 | 1303 | 2 B84493 | probable Athila re |
| 7 | 79.5 | 12.2 | 507 | 2 S64507 | bone morphogenetic |
| 8 | 79.5 | 11.8 | 1038 | 2 JC5527 | PAR interacting pr |
| 9 | 79.5 | 11.8 | 1277 | 2 T32731 | hypothetical prote |
| 10 | 78.5 | 11.6 | 264 | 2 T23054 | hypothetical prote |
| 11 | 78.5 | 11.6 | 550 | 2 T26562 | two-component sens |
| 12 | 78 | 11.6 | 682 | 2 B84021 | hypothetical prote |
| 13 | 78 | 11.6 | 797 | 2 A36811 | hypothetical prote |
| 14 | 77.5 | 11.5 | 191 | 2 T15129 | hypothetical prote |

| | | | | | |
|----|------|------|------|----------|--------------------|
| 15 | 77 | 11.4 | 555 | 2 G90594 | hypothetical prote |
| 16 | 76.5 | 11.4 | 168 | 2 A96654 | hypothetical prote |
| 17 | 76.5 | 11.4 | 786 | 2 C86406 | 88.6k hypothetical |
| 18 | 76 | 11.3 | 559 | 2 A53423 | TPCCT-gamma protel |
| 19 | 76 | 11.3 | 1255 | 2 B97104 | hypothetical prote |
| 20 | 76 | 11.3 | 1278 | 2 B70236 | hypothetical prote |
| 21 | 75.5 | 11.2 | 336 | 2 T08538 | hypothetical prote |
| 22 | 75.5 | 11.2 | 491 | 2 T16354 | hypothetical prote |
| 23 | 75 | 11.5 | 344 | 2 T40167 | hypothetical prote |
| 24 | 75 | 11.1 | 360 | 2 T34510 | hypothetical prote |
| 25 | 75 | 11.5 | 534 | 2 T33903 | serine-rich protel |
| 26 | 75 | 11.1 | 2410 | 2 T43731 | cell wall alpha-g1 |
| 27 | 75 | 11.1 | 3844 | 2 T18402 | asparagine/aspara |
| 28 | 74.5 | 11.1 | 146 | 1 PSDG | phospholipase A2 (|
| 29 | 74.5 | 11.1 | 146 | 1 PSEGA | phospholipase A2 (|
| 30 | 74.5 | 11.4 | 328 | 2 S67570 | hypothetical prote |
| 31 | 74.5 | 11.1 | 643 | 2 T21428 | hypothetical prote |
| 32 | 74 | 11.0 | 296 | 2 D70465 | hypothetical prote |
| 33 | 74 | 11.4 | 600 | 2 S07638 | spore coat protein |
| 34 | 74 | 11.0 | 612 | 2 E84809 | hypothetical prote |
| 35 | 74 | 11.0 | 4466 | 1 S17231 | dynein beta heavy |
| 36 | 73.5 | 10.9 | 208 | 2 T06492 | ribosomal protein |
| 37 | 73.5 | 11.3 | 222 | 2 E81564 | conserved hypothet |
| 38 | 73.5 | 11.3 | 228 | 2 H72103 | ypdp hypothetical |
| 39 | 73.5 | 11.3 | 228 | 2 A86518 | hypothetical prote |
| 40 | 73.5 | 10.9 | 357 | 2 T34012 | hypothetical prote |
| 41 | 73.5 | 10.9 | 562 | 2 B41035 | chitinase (EC 3.2. |
| 42 | 73.5 | 10.9 | 562 | 2 S50371 | chitinase (EC 3.2. |
| 43 | 73.5 | 11.3 | 636 | 2 S63131 | probable membrane |
| 44 | 73.5 | 10.9 | 2339 | 2 A45597 | DNA-directed RNA p |
| 45 | 73.5 | 10.9 | 4572 | 2 S57908 | hypothetical 527k |

ALIGNMENTS

RESULT 1
A42435
leech antiplatelet protein precursor - Mexican leech
C:Species: Haementeria officinalis (Mexican leech)
C:Date: 07-Apr-1994 #sequence: revision 07-Apr-1994 #text: change 07-May-1999
C:Accession: A42435; A42434
R:Keller, P.M.; Schultz, L.D.; Condra, C.; Karczewski, J.; Connolly, T.M.
J. Biol. Chem. 267, 6899-6904, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42435; MUID:92202247; PMID:1551898
A:Accession: A42435
A:Molecule type: mRNA
A:Residues: 1-147 <XEL>
A:Cross-references: GB:M81489
A>Note: the authors' translation is shown at residue 65
A>Note: the authors' translated the codon AGC for residue 20 as Arg and GAA for residu
R:Connolly, T.M.; Jacobs, J.W.; Condra, C.
J. Biol. Chem. 267, 6893-6898, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42434; MUID:92202246; PMID:1551897
A:Accession: A42434
A:Molecule type: protein
A:Residues: 59-91;101-113;123-139 <CON>
F:1-21/Domains: signal sequence #status predicted <SIG>
F:12-147/Product: leech antiplatelet protein #status predicted <MAN>
Alignment Scores:
Pred. No.: 0.366
Score: 88.50
Percent Similarity: 35.25%
Best Local Similarity: 24.59%
Query Match: 13.13%
DB: 2
Gaps: 4
Length: 147
Matches: 30
Conservative: 13
Mismatch: 32
Indels: 47
US-09-936-737A-1 (1-375) x A42435 (1-147)
QY 1 ATGAAGTATTCTGATTTCTCTTGCCTTCGCGAAGCTTGATGATCTCA----- 51

```

Db 1 MetasSerPheLeuPheSerLeuAlaLys-----SerLeuValAlaIleProAla 18
OY 51 -----
Db 19 IleSerAlaGlnAspGlnAspAlaGlyValaGlyAspGlnIleSerGluGlyGluAsp 38
OY 52 ACTACTTCTTCAGAGAA-----
Db 39 ThrThrGlySerAspGluThrProSerThrGlyGlyGlyAspGlyValAsnGluGlu 58
OY 70 -----CGTGAAGATTGTTGGACGTTTACCGCAACAGAAATATACAGAC 114
Db 59 ThrIleThrAlaGlyAsnGlyAspCysTrpSerIysArgProGlyTrpIysLeuProAsp 78
OY 115 TTCGATTAATGCTTTAAGACGCTTCGATTCGACGAAATGCAAAAAACATGTTTACG 174
Db 79 AsnLeuLeuThrLysThrGluPheThrSerValAspGlyCysArgLysMetCysGluGlu 98
OY 175 ACGGATGAC-----TGCTACATCGTTTAAAGACACGCTCAACAGAAATGTTTAC 225
Db 99 SerAlaValaGluProSerCysTrpIleLeuGlnIleAsnThrGluThrAsnGluCysTyr 118
OY 226 TACAAAT 231
Db 119 ArgAsn 120

```

RESULT 2

hypothetical protein F54D12.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T32824

R:Magdi, L.; Goela, D.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F54D12.

A:Reference number: Z21228

A:Accession: T32824

A:Status: preliminary: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-164 <MAG>

A:Cross-references: EMBL:AF040647; PIDN:AAB94992.1; GSPDB:GN00020; CESP:F54D12.1

A:Experimental source: strain Bristol N2; clone F54D12

A:Gene: CESP:F54D12.1

A:Map position: 2

A:Introns: 91/3; 138/1

Alignment Scores:

```

Pred. No.: 0.975      Length: 164
Score: 84.50      Matches: 23
Percent Similarity: 45.35%      Conservative: 16
Best Local Similarity: 26.74%      Mismatches: 34
Query Match: 12.54%      Indels: 13
Db: 2      Gaps: 3

```

US-09-936-737A-1 (1-375) x T32824 (1-164)

```

OY 10 TTCTGATTTCCTTCCTTCCTTCGCAAGCTGCTGATCTCACTACTTCTTCAGAGAA 69
Db 5 PheLeuIleCysTrpValLeuValAsnSerIleMetValSerValArgValAlaProGlu 24
OY 70 CGTGAAGATTGTTGGACGTTTACCGCAACAGAAATATACAGCTTCGATTAATCTTTT 129
Db 25 IleSer-----ThrSerTrpSerThrGluIle 33
OY 130 AAGAAGTCTCTGATCTTGAACGAATGCAAAAAACATGTTTCAAGACGAGTCTGCTAC 189
Db 34 IleAsnSerLeuThrTrpAspGluCysValIleGlnCysLeuSerThrGluIleCysIle 53
OY 190 ATCGTTTTTGAAGACAGGTCACAAAGAAATGTTAC---TACATGTCGTTGATGGTAA 246
Db 54 MetAlaTrpSerAsnSerLeuAsn---IleCysTrpLeuThrValAlaValGlyAspValIle 72

```

```

OY 247 GAGTAGACCAAGAAAA 264
Db 73 GluValArgHisAspGln 78

```

RESULT 3

A55320

immunophilin FKBP46 - fall armyworm

C:Species: *Spodoptera frugiperda* (fall armyworm)

C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 02-Sep-2000

C:Accession: A55320

R:Alnemri, E.S.; Fernandes-Alnemri, T.; Pomeroy, K.; Robertson, N.M.; Dudley, K.; D

J. Biol. Chem. 269, 30828-30834, 1994

A:Reference number: A55320; PMID:95074110; PMID:7527037

A:Accession: A55320

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-412 <ALIN>

A:Cross-references: GB:U15038; NID:9595844; PID:9595845

C:Superfamily: Yeast peptidylprolyl isomerase FR3; BKB-type peptidylprolyl isomerases

C:Keywords: nucleus; phosphoprotein

F;324-371/Domain: BKB-type peptidylprolyl isomerase homology <PPI>

Alignment Scores:

```

Pred. No.: 1.67      Length: 412
Score: 82.50      Matches: 32
Percent Similarity: 41.74%      Conservative: 16
Best Local Similarity: 27.83%      Mismatches: 40
Query Match: 12.24%      Indels: 27
Db: 2      Gaps: 5

```

US-09-936-737A-1 (1-375) x A55320 (1-412)

```

OY 43 CTGATCTCAACTACTTCTTCACAGACGTAAGATTTGGACGTTTACCGCAACAGA 102
Db 116 LeuValProAlaLysAsnLysArgLysLeuGlnAsnAlaAsnAspAlaThrAsnLys 135
OY 103 AATATACAGACTTCATTAATCTTTTAGAGTCC-----TCTGATCTT 147
Db 136 LysAlaLysProAspLysLysAlaGlyLysAsnSerAlaProAlaAlaLysSerAsp 155
OY 148 GAGCAATGCAAAAAAACATGTTTCAAGACGAGTACTCTACATCGTTTGAAGACAG 207
Db 156 AspAspAspAsp-----GlnAspGln 162
OY 208 GTCAACAGAAATGTTTACTACAAATGCTGATGTTGAAGATTAACCAAGAAATTT 267
Db 163 LeuGlnLys-----PheLeuAspGlyGluAspIleAspThrAspGlu--- 176
OY 268 GTTGTGACGAAAACTTCACGAAATTTTTCGACAGACTCGGAGGTAAAGATGACAGT 327
Db 177 ---AsnAspGluSerPheLysMetAsn-----ThrSerAlaGluGlyAspAspSerAsp 193
OY 328 AATGCGCGAGTACAGGTGACGATCAGATGATGATGATGATGATGATGATGATGAT 372
Db 194 GluGluAspAspAspGluAspGluGluAspGluGluAspAspAsp 208

```

RESULT 4

T00548

hypothetical protein At2g39380 (imported) - *Arabidopsis thaliana*

N:Alternate names: hypothetical protein F12L6.4

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001

C:Accession: T00548; F84816

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, July 1998

A:Description: *Arabidopsis thaliana* chromosome II BAC F12L6 genomic sequence.

A:Reference number: Z14168

A:Accession: T00548

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-637 <ROU>


```

Db      166 MetLeuLeuAspThrThrSerAsnGlyAsnPhenLeuAsnLysHisValGluGluGlyTyr 185
QY      85 ACCTTTACGACGACGAGAAATATACA-----GACTTCGATTAATCTC 126
Db      186 GluLeuValGluAsnLeuAlaGlnSerAspLysAsnGlyAsnLysAspTyrGluAsnSer 205
QY      127 TTTAAGAGATCCTCTGATCTTGACGAATGCAAAAATCATGT-----TTCAAGACGGAG 180
Db      206 IleArgThrSerSerAspSerAspGlu-----LysHisCysArgGluMetLysLysHis 223
QY      181 TACTCTACATCGTTTGTGAAGACACGGCTCAACAAGAAATGTACTACATGTCGTGAT 240
Db      224 ValHisPheLeuGlyAspAspGluThr-----PheGlnValGlnAsp 237
QY      241 GGTGAAGATTAGACCAAGAAATTTGTTGTGACGAAACTCTACGGAATATTATTTG 300
Db      238 GlyGluThrLeuGlnSerGluLysVal-----AsnTyrVal 249
QY      301 ACAGACGCGGAGGTAAGATGCAAGTAAATCGCGGACGATGACGAG 351
Db      250 GlnAsnHisGlyTyrAsnLysGlyLysSerIleGlnThrProLysGlu 266

```

RESULT 7

Probable membrane protein YGR189c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein G7553
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
 R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64499
 A:Accession: S64507
 A:Molecule type: DNA
 A:Residues: 1-507 <ARR>
 A:Cross-references: EMBL:Z72974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN00007;
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:CRM1; MIPS:YGR189C
 A:Cross-references: SGD:S0003421
 A:Map position: 7R
 C:Keywords: transmembrane protein
 F:6-22/Domain: transmembrane #status predicted <TM>

Alignment Scores:

| Pred. No.: | 3.5 | Length: | 507 |
|------------------------|--------|---------------|-----|
| Score: | 79.50 | Matches: | 42 |
| Percent Similarity: | 43.18% | Conservative: | 15 |
| Best Local Similarity: | 31.82% | Mismatches: | 50 |
| Query Match: | 12.21% | Indels: | 25 |
| | | Gaps: | 4 |

US-09-936-737a-1 (1-375) x S64507 (1-507)

```

QY      371 TCTTCATCAACT-----TCATCTGACCTGTCACCTGACCTGCCGCAATCTTCGACATCT 318
Db      308 SerSerSerThrValSerSerSerAlaSerSerThrValSerSerSerValSerSerThr 327
QY      317 TTACCCCGACGCTGCTCAATAATTTTCCTGGAAGTTTGTGACACAAACAATTTTCT 258
Db      328 ValSerSerSerAlaSerSerThrValSerSerSerValSerSerThrValSerSerSer 347
QY      257 TGTCTTAAGCTTTACCATCAAGACATTTGTAAGTTCCTCTGTTGACCGGTCTTCA 198
Db      348 SerSerValSerSerSerSerSerThrSerProSerSerSerThrAlaThr-----SerSer 366
QY      197 AAACAGATGTACAGTACTCCGCTTG----- 171
Db      367 LysThrLeuAlaSerSerSerValThrThrSerSerSerIleSerSerPheGluLysGln 386
QY      170 -----AAACATGTTTTCATTCGTCACAGATCAAGAGACTTCTTAAAA 126

```

```

Db      387 SerSerSerSerSerLysThrValAlaSerSerSerThrSerGluSerIleLeuSer 406
QY      125 GAT-----TTATGAGAGCTGTATATTTTTCGTCGCGTAAACGTC 84
Db      407 SerThrLysThrProAlaThrValSerSerThrThrArgSerThrValAlaProThrThr 426
QY      83 CACCAATCTTCACGCTTCTTCTTCAAGAACTACTGAG 48
Db      427 GlnGlnSerSerValSerSerSerProValGln 438

```

RESULT 8

JC5527
 bone morphogenetic protein type II receptor precursor - mouse
 C:Species: *Mus musculus* (house mouse)
 C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 24-Sep-1999
 C:Accession: JC5527
 R:Beppu, H.; Minowa, O.; Miyazono, K.; Kawabata, M.
 Blochem. Biophys. Res. Commun. 235, 499-504, 1997
 A:Title: cDNA Cloning and genomic organization of the mouse BMP type II receptor.
 A:Reference number: JC5527; MUID:97350808; PMID:9207184
 A:Accession: JC5527
 A:Molecule type: mRNA
 A:Residues: 1-1038 <REP>
 A:Cross-references: DDB:AF003942; NID:92253704; PIDN:AA63042.1; PID:92253705
 C:Comment: This protein is a serine/threonine kinase receptor that forms a heteromeric
 C:Genetics:
 A:Gene: BMPR-II
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:151-174/Domain: transmembrane #status predicted <TM>
 F:201-508/Domain: protein kinase homology <KIN>
 F:202-500/Domain: serine/threonine kinase #status predicted <STK>

Alignment Scores:

| Pred. No.: | 3.65 | Length: | 1038 |
|------------------------|--------|---------------|------|
| Score: | 79.50 | Matches: | 27 |
| Percent Similarity: | 40.46% | Conservative: | 26 |
| Best Local Similarity: | 20.61% | Mismatches: | 35 |
| Query Match: | 11.80% | Indels: | 43 |
| | | Gaps: | 4 |

US-09-936-737a-1 (1-375) x JC5527 (1-1038)

```

QY      10 TCTTGATTCCTTCCTTCGCTGCAAGCTCTGTAATCAACATCTTTCAGAGAA 69
Db      9 PheArgValProThrPheLeuThrPheValLeuValSerThrThrAlaAlaSerGln 28
QY      70 CCGTAAGATTTGTGACGCTTTTACGCGACGAAATATATACAGACTTTCGATTAATCTTT 129
Db      29 AsnGlnGln-----ArgLeuCysAlaPheLysAspProTyr 40
QY      130 AAGAGTCTCTGATCTTGACGAATGCAAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyGlnSerArgIleSerHisGluAsnGlyThrIleLeuLys 60
QY      169 TTCAAGACGAGATCTCTACATCGTTTTCGAA-----GACACGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTyrGlyLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAGAAATGTTACTCAATGTCGTTGATGTAAGAGATTAAGCAAGAAATTTGTT 270
Db      81 LysGlnGlySerThrSerHisIleGlyAspProGlnGlyCysHisTyrGluGlnCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrGlyArgPheCysCysSerThr 120
QY      274 -----GACCAAACTTCACGAGAAATTAAT 297
Db      121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131

```

RESULT 9

T32731

PAR interacting protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T32731
R:Comte, P.A.; Ossipow, V.; Schibler, U.
Submitted to the EMBL Data Library, January 1997
A:Description: Isolation of PIP, a 160 kDa nuclear protein that interacts with the act
A:Reference number: 221213
A:Accession: T32731
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1277 <COM>
A:Cross-references: EMBL:083590; NID:92253210; PID:92253211; PIDN:AMB62878.1

Alignment Scores:

| Pred. No.: | 3.69 | Length: | 1277 |
|------------------------|--------|---------------|------|
| Score: | 79.50 | Matches: | 33 |
| Percent Similarity: | 40.87% | Conservative: | 14 |
| Best Local Similarity: | 28.70% | Mismatches: | 43 |
| Query Match: | 11.80% | Indels: | 25 |
| DB: | 2 | Gaps: | 5 |

US-09-936-737A-1 (1-375) x T32731 (1-1277)

```

Oy 25 CTTTCCTTCGCAAGCTTCTGTCACACTCTTCAGAGAAGCTGAGATTGTTGG 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 LeuGlnLeuIleLeuAlaValLeuAsnProGluThrAsnGluAspGluAsp----- 634
Oy 85 AGCTTTACGCGACAGAAATATACAGACTTCGAT---AAATCTTTAACAATCTCT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 -----AsnValValValThrAspThrAspGluLysGlnLeuLysHisGlyGlu 650
Oy 142 GATCTTGACGATGCAAAAAAATGTTTCAGACGAGTACTGCTACATGTTTGA 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 AspAlaAspSerAspSer-----GluVal 657
Oy 202 GACACGCTTCAACAAGATGTTACTACATGTCGTTGATGTGAGAGTTAGACCAAGAA 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 AspSerLysAsnSerGluSerAspValAspSerGluAspGluSerGluGluGlu 677
Oy 262 -----AAATTTGTTGCGACGAAACTTCACGCGAAATATATTGTCAGACTGCGAGGT 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 AspArgAspLysAspValAspProGlyPheArgGlnGlnLeuMet-----GluVal 694
Oy 316 AAAATGACAGTATGCGCGACGATACAGGTGACAGTGCAGATGAA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 LeuGlnAlaGlyAsnAlaLeuGlyGlyGluGluGluGluGlu 709

```

RESULT 10

T22054
hypothetical protein F40G12.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000

C:Accession: T22054

R:McMurray, A.
submitted to the EMBL Data Library, July 1996

A:Reference number: 219507
A:Accession: T22054

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-264 <WIL>
A:Cross-references: EMBL:Z77661; PIDN:CAB01183.1; GSPDB:GN00023; CESP:F40G12.5

A:Experimental source: clone F40G12

C:Genetics:

A:Gene: CESP:F40G12.5

A:Map position:

A:Introns: 57/3: 158/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F40G12.4

Alignment Scores:

| Pred. No.: | 4.31 | Length: | 264 |
|---------------------|--------|---------------|-----|
| Score: | 78.50 | Matches: | 35 |
| Percent Similarity: | 41.54% | Conservative: | 19 |

| Best Local Similarity: | 26.92% | Mismatches: | 47 |
|------------------------|--------|-------------|----|
| Query Match: | 11.65% | Indels: | 29 |
| DB: | 2 | Gaps: | 8 |

US-09-936-737A-1 (1-375) x T22054 (1-264)

```

Oy 10 TTTCTGATTTCTCTTCCTTTGCTTCGACAGCTTCGATCTCAACTACTCTTCAGAGAA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 PheAlaIleSerAlaAsnLysSerIleLeuSerPheHisArgAsnLeuLeuGluGlu 30
Oy 70 CGTGAAGATGTTGAGCGTTTACGCGAAGAAATATACAGACTTCGATTAATCTTT 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 SerGluAspCys-----PheGlnLysValPhe 39
Oy 130 AAGAG--TCCTGTGATCTTCGACGAATGCAAAAAAATGTTTCAAGAGAGTCTGC 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40 LeuAlaIleIleSerGlyLysHisGlyLysSerLys-----AspTyrAsp 54
Oy 187 TACATCGTTTTCGAA-----GACACGCTCAACAAGAAATGTTTC 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 PheLeuAlaArgAsnLeuIleGlnArgGluAlaLeuThrSerGlyLysGlyCysPhe 74
Oy 226 TACAAATGCTGTTGATGTGAGAGACTTAGACCAAGAAATTT--GTTTCGACGAGAAAC 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 LeuGlnIleValLys--GluGluCysProGluLysPheLysLeuIleGluGluAsn 93
Oy 283 TTCACGGA--AAATATTGACAGACTGCGAGGTAAGATGCAAGTAAATGCGCA 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 TyrSerGlnLeuValThrLeuLeuThrGlu--LysProLysAspAsnGlyAlaCysThr 112
Oy 337 GTTACAGCTGACGAGTCAAGATGATGAT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 AlaProTyrPheGlnLeuGluGluGlu 122

```

RESULT 11

T26562
hypothetical protein Y26D4A.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26562

R:White, S.
submitted to the EMBL Data Library, September 1999

A:Reference number: Z20234
A:Accession: T26562

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-550 <WIL>
A:Cross-references: EMBL:AL110478; NID:e1542139; PIDN:CAB54340.1; CESP:Y26D4A.2

A:Experimental source: clone Y26D4A

C:Genetics:

A:Gene: CESP:Y26D4A.2
A:Introns: 134/3: 160/1; 286/3; 316/1; 480/2; 502/2

Alignment Scores:

| Pred. No.: | 4.49 | Length: | 550 |
|------------------------|--------|---------------|-----|
| Score: | 78.50 | Matches: | 23 |
| Percent Similarity: | 38.89% | Conservative: | 12 |
| Best Local Similarity: | 25.56% | Mismatches: | 32 |
| Query Match: | 11.65% | Indels: | 23 |
| DB: | 2 | Gaps: | 3 |

US-09-936-737A-1 (1-375) x T26562 (1-550)

```

Oy 79 TGTGGACGTTTACGCGACAGAAATATACAGACTTC----- 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 CysGlyAsnPheLeuHisAsnArgProTyrProAspSerCysGluThrGlyCysMetAsp 189
Oy 118 -----GATAAATCTTTAAGAG 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 ThrMetValGlnValAsnAlaLysProGlyProLeuMetGlySerArgAsnIleLysAsp 209
Oy 136 TCC--TCTGACCTTACCAATGCAAAAAAATGTTTCAAGACGAGTACTGCTACATC 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 210 AsnLeuThrSerSerAspGluCysValIleuTyrCysTrpIysAspLeuAsnCysPheVal 229
 QY 153 GTTTTGAAGACACGGCTAACACAGAAATGTTACTACAATGCGTTGATGGTGAAGAGTTA 252
 Db 230 AsnTrpTrpAspLysAspSerLysGluCysTrpTrpPheIleAspAsnValHisPhe 249
 QY 253 GACCAAGAAAAATTTGTTGTCGACGAAAC 282
 Db 250 --LeuGluLysValHisProSerGluAsn 258

RESULT 12

B84021

two-component sensor histidine kinase involved in chemotaxis cheA [Imported] - Bacillus

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B84021

R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirata

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84021

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-682 <STO>

A:Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA06689.1; GSPDB:GNOC

A:Experimental source: strain C-125

A:Gene: cheA

C:Genetics:

C:Superfamily: chemotaxis protein cheA

Alignment Scores:

Pred. No.: 5.13 Length: 682

Score: 78.00 Matches: 28

Percent Similarity: 47.67% Conservative: 13

Best Local Similarity: 32.56% Mismatches: 31

Query Match: 11.57% Indels: 14

DB: 2 Gaps: 5

US-09-936-737a-1 (1-375) x B84021 (1-682)

QY 91 TACGGCAAGAAATATACAGACTTGCATTAATCTTTAAGAATGCTCTGAT----- 144
 Db 156 TyrAspGluPheGluMetThrValLeuAspGlnSerPheGluGlnGlyTrpGlnAlaTyr 175
 QY 145 -----CTTGACGAATGCMAAAACAAATGTTTCAAGACGGAGTACTGCTAC 189
 Db 176 GlnIleGluValThrLeuAspGlu-----LysThrLeuLeuLysAlaAlaArgValPhe 193
 QY 190 ATCGTTTGTGA-----GACACGGTCAACACAGATGTTACTACATGCTGTGATGT 243
 Db 194 MetValPheGluValLeuGluGlnValGly---GluValIleLysSerThrProSerAla 212
 QY 244 GAGAGTTTACCAACAAATTTGTTGTCAGCAAAACCTTCACGCAAAATATATTGACA 303
 Db 213 GlnGluLeuGlnGluLysPhe-----AspGluArgPheLeuValThrLeuLeuThr 230
 QY 304 GACTCGAGAGGTAAAGAT 321
 Db 231 LysValAspGlyGluGlu 236

RESULT 13

A36811

hypothetical protein ORF48 - saimiriine herpesvirus 1 (strain 11)

C:Species: saimiriine herpesvirus 1

A:Note: host Saimiri sciureus (common squirrel monkey)

C>Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999

C:Accession: A36811

R: Albrecht, J.

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A36806

A:Accession: A36811

A:Molecule type: DNA

A:Residues: 1-797 <ALB>
 A:Cross-references: GB:X64346; NID:960320; PIDN:CAA45671.1; PID:960369
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Blesinger, B.; Newman, C.;
 J. Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A37309; MUID:9233688; PMID:1321287
 A:Contents: annotation; protein-coding frames
 A:Note: neither protein nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 48

Alignment Scores:

Pred. No.: 5.17 Length: 797

Score: 78.00 Matches: 31

Percent Similarity: 44.79% Conservative: 12

Best Local Similarity: 32.29% Mismatches: 33

Query Match: 11.57% Indels: 20

DB: 2 Gaps: 5

US-09-936-737a-1 (1-375) x A36811 (1-797)

QY 94 GCGAAGAAATATACAGACTTGCATTAATCTTTAAGAATGCTCTGATCTGACGAA 153
 Db 382 AlaAsnGluLys-----GluTrpLysLysIleIleAspLysSerAspAspArgAspAsp 399
 QY 154 TGCMAAAACAAATGTTTCAAGACGGAGTACTGATCGTTTGAAGACAGGTCAC 213
 Db 400 ArgAspLysAspGluTrpGluLys-----Asn 409
 QY 214 AAGAAATGTTACTACAAAT-----GTGTTGATGTTGAAGATGACCAAGAA 261
 Db 410 GluGlu-----TyrAsnArgAspGluGluGluAspGluGluGluAspGluGlu 427
 QY 262 AAATTTGTTGTCGACGAAATCTTCACGCAAAATATTTGACACACGCCAGGTTAAGAT 321
 Db 428 LysAspGluLysGluGluGluGluGluAspGluGluAspGluGluGluGluAsp 447
 QY 322 GCAGTAAATGCGGACGATACAGTGCAGCTCATGAAATGATGAA 369
 Db 448 GluGluGlu-----AspGluGluLysAspGluGluGluGluGluGluGlu 461

RESULT 14

T15129

hypothetical protein ZC328.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000

C:Accession: T15129

R: Mamsley, P.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid ZC328.

A:Reference number: Z18298

A:Accession: T15129

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <WAM>

A:Cross-references: EMBL:AF000194; NID:G1946990; PID:G1946994; PIDN:AAB52895.1; GSPDB

A:Experimental source: strain Bristol N2; clone ZC328

C:Genetics:

A:Gene: CESP:ZC328.5

A:Map position: 1

A:Introns: 20/3; 124/3

C:Superfamily: Caenorhabditis elegans hypothetical protein ZC328.5

Alignment Scores:

Pred. No.: 5.4 Length: 191

Score: 77.50 Matches: 23

Percent Similarity: 35.51% Conservative: 15

Best Local Similarity: 21.50% Mismatches: 28

Query Match: 11.50% Indels: 41

DB: 2 Gaps: 4

US-09-936-737a-1 (1-375) x T15129 (1-191)

| | | | |
|----|-----|--|-----|
| OY | 241 | - - - - GGTAAAGATTGACCAAGAATAATTTTCTTCGCACAATACTTCACGGAAAT | 294 |
| Dd | 153 | GlnlyserGelncInlIleasnTrpArgSerPheThrPeLuvalllysasProalaspn | 17Z |
| OY | 295 | TATTTGACACTCGCAGGGAAGATGTCAGGTAAAGCCGCCAGTGTCAGGTACACGTCA | 354 |
| Dd | 173 | IleargPhaeapProcIlulyslleaspTyraAspnalaLyseLuileelyLysteutls | 19Z |
| OY | 355 | GATGAAGTTGAT | 366 |
| Dd | 193 | Asnlystleasp | 196 |

```
Search completed: February 25, 2003, 02:29:12
Job time : 27 secs
```

193 ASnLYStleasp 196

eAsp 196

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic acid - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:50:35 ; Search time 12.5 Seconds

(Without alignments)
2488.580 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 674

Sequence: 1 atgaagtattcttgattc.....atgaagtattctgaatataa 375

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_n2p
-O=/cgm2.1/USPTO/spool/US09936737/runtat_14022003_100557_1389/app_query.fasta.1.519
-DB=SwissProt_40 -PMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdt -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09936737@cgm2.1.1.6 @runtat_14022003_100557_1389 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGESIZE=0 -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEROUT=120
-NARN_TIMEROUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 1 | 95.5 | 14.2 | 147 | 1 | LAPP_HAEOP |
| 2 | 82.5 | 12.2 | 412 | 1 | FKB1_SPOFR |
| 3 | 79.5 | 12.2 | 507 | 1 | YGA6_YEAST |
| 4 | 79.5 | 11.8 | 1038 | 1 | BMR2_MOUSE |
| 5 | 78 | 11.6 | 797 | 1 | VG48_HSVSA |
| 6 | 76 | 11.3 | 559 | 1 | TCRG_TETPY |
| 7 | 75.5 | 11.2 | 419 | 1 | ICBC_MOUSE |
| 8 | 75.5 | 11.2 | 491 | 1 | ICBC_MOUSE |
| 9 | 75 | 11.1 | 360 | 1 | YORA_CAEEL |
| 10 | 75 | 11.1 | 2410 | 1 | MOK1_SCHPO |
| 11 | 74.5 | 11.1 | 146 | 1 | PA21_CANPA |
| 12 | 74.5 | 11.1 | 146 | 1 | PA21_CANPA |
| 13 | 74.5 | 11.1 | 451 | 1 | PA21_CANPA |
| 14 | 74 | 11.0 | 296 | 1 | Y723_HUMAN |
| 15 | 74 | 11.0 | 600 | 1 | SP96_DIQAE |
| 16 | 73.5 | 10.9 | 4465 | 1 | DYHC_ANTCR |
| 17 | 73.5 | 10.9 | 208 | 1 | RK1_PEA |
| 18 | 73.5 | 10.9 | 562 | 1 | CHIT_YEAST |

| | | | | | | |
|---|----|------|------|------|---|------------|
| C | 19 | 73.5 | 11.3 | 636 | 1 | YNR6_YEAST |
| C | 20 | 73.5 | 10.9 | 2339 | 1 | RPCL_PLAFA |
| C | 21 | 73 | 11.2 | 551 | 1 | YGF1_YEAST |
| C | 22 | 73 | 10.8 | 1405 | 1 | DPOA_SCHPO |
| C | 23 | 72.5 | 11.1 | 1802 | 1 | HKR1_YEAST |
| C | 24 | 72 | 11.1 | 338 | 1 | YD49_THENA |
| C | 25 | 72 | 10.7 | 587 | 1 | YKRO_YEAST |
| C | 26 | 71.5 | 10.6 | 146 | 1 | PA21_RAT |
| C | 27 | 71.5 | 10.6 | 156 | 1 | CB20_HUMAN |
| C | 28 | 71.5 | 10.6 | 1177 | 1 | Y307_MYCE |
| C | 29 | 71 | 10.9 | 110 | 1 | YCT1_YEAST |
| C | 30 | 71 | 10.9 | 373 | 1 | YVU2_YEAST |
| C | 31 | 71 | 10.5 | 782 | 1 | YVU2_YEAST |
| C | 32 | 71 | 10.5 | 2555 | 1 | FAFY_HUMAN |
| C | 33 | 70.5 | 10.5 | 78 | 1 | MOH1_CANGG |
| C | 34 | 70.5 | 10.5 | 346 | 1 | UFEL_YEAST |
| C | 35 | 70.5 | 10.5 | 748 | 1 | Y875_MENJA |
| C | 36 | 70.5 | 10.5 | 793 | 1 | NETB_DROME |
| C | 37 | 70 | 10.4 | 763 | 1 | MDOB_ECOLI |
| C | 38 | 70 | 10.4 | 1058 | 1 | S185_YEAST |
| C | 39 | 69.5 | 10.7 | 264 | 1 | UL31_HSV6U |
| C | 40 | 69.5 | 10.3 | 733 | 1 | YU08_YEAST |
| C | 41 | 69.5 | 10.7 | 1468 | 1 | N153_RAT |
| C | 42 | 69 | 10.2 | 644 | 1 | KNG_HUMAN |
| C | 43 | 69 | 10.2 | 1056 | 1 | RIC1_YEAST |
| C | 44 | 68.5 | 10.2 | 591 | 1 | CAIX_RAT |
| C | 45 | 68.5 | 10.2 | 610 | 1 | CAIX_HUMAN |

ALIGNMENTS

RESULT 1
LAPP_HAEOP
ID LAPP_HAEOP STANDARD: PRT: 147 AA.

AC Q01747;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Anti-platelet protein precursor.

GN LAPP.

OS Haementeria officinalis (Mexican leech).

OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
OX NCBI_Taxid=6410; Glossiphoniidae; Haementeria.

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.
RC TISSUE=Salivary gland;
RX MEDLINE=92202247; PubMed=1551898;

RA Keller P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;
RT salivary glands of collagen-stimulated platelet activation from the
RT the cDNA and expression.

RL J. Biol. Chem. 267:6899-6904(1992).

CC - FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET
CC AGGREGATION, DENSE GRANULE RELEASE AND SEROTONIN RELEASE.
CC - SUBCELLULAR LOCATION: Secreted

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC DR EMBL; M81489; AAA29194.1; -

CC KW Signal.

CC FT SIGNAL

CC CHAIN 1 21 POTENTIAL.

CC SEQUENCE 147 AA; 15908 MW; 75A5511374AAE42E CMC64;

CC Alignment Scores: 0.0379 Length: 147

Pred. No.:

```

Score: 95.50 Matches: 31
Percent Similarity: 36.07% Conservative: 13
Best Local Similarity: 25.41% Mismatches: 31
Query Match: 14.17% Indels: 47
DB: 1 Gaps: 4

US-09-936-737A-1 (1-375) x LAMP_HAE0F (1-147)

OY 1 ATGAGATATTTCGATTTCCTTCCTTCGCGACGAGCTGCATGCTCA----- 51
    ||| ||||| ||| ||| |||||:::
DB 1 MetasnerphenleuphenSerleuAlaCys-----SerleuValAlaIleproAla 18
OY 51 ----- 51
DB 19 IleSerAlaGlnAspGluAspAlaGlyGlyAlaGlyAspGluThrSerGluGlyLys 38
OY 52 ACTACTCTTACAGAA----- 69
    ||||| ||:::|
DB 39 ThrThrGlySerAspGluThrProSerThrGlyGlyGlyAspGlyLysGlu 58
OY 70 -----CGTGAAGATTGTGGACGTTTACGCGACAGCAAAATATACAGC 114
    ||||| ||:::|
DB 59 ThrIleThrAlaGlyAsnGluAspGlyTyrSerIleSarGProGlyThrIlePheLeuProAsp 78
OY 115 TTCGATAAATCTTTTAAAGAGTCTCTGCATCTTGACGATGCAAAAAACATGTTCAAG 174
    ::: ::::: |||||:::| || :::
DB 79 AsnIleuThrIleThrGluPheThrSerValAspGluSerIleSarGlySmetGlyGlu 98
OY 175 ACGAGATAC-----TGCTACATCGTTTGAAGACACGCGTCAACAGATGTTAC 225
    ::: ::::: |||||:::|
DB 99 SerAlaValAlaGluProSerCysIleThrIleuGlnIleAsnThrGluThrAsnGluCysTyr 118
OY 226 TACAAAT 231
    |||
DB 119 ArgAsn 120

RESULT 2
FKB4_SPOFR STANDARD: PRT: 412 AA.
ID FKB4_SPOFR
AC Q26486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
DE isomerase) (PPIase) (EC 5.2.1.8).
OS Spodoptera frugiperda (Fall armyworm).
OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuoidea; Noctuidae; Amphyrinae; Spodoptera.
CX NCBI_TaxId=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95074110; Pubmed=7527037;
RA Aluemi E.S., Fernandes-Aluemi T., Pomeroy K., Robertson N.M.,
RA Dudley K., Dubois G.C., Litwack G.,
RT "FKBP46, a novel Sf9 insect cell nuclear immunophilin that forms a
RT protein-kinase complex."
RL J. Biol. Chem. 269:30828-30834(1994).
RC
CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS
CC DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+
CC AND ATP.
CC
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

| Accession | Protein Name | Length (aa) | MW (kDa) | pI (pI) | Score | Similarity (%) | Local Similarity (%) | Query Match (%) | Gaps |
|--|--|---------------|----------|---------|-------|----------------|----------------------|-----------------|------|
| CC | or send an email to license@sb-ib.ch . | | | | | | | | |
| CC | EMBL; U15038; AAA58962.1; - | | | | | | | | |
| DR | HSSP; P20077.1; IFKJ | | | | | | | | |
| DR | InterPro; IPR001179; FKBP_PPIase. | | | | | | | | |
| DR | Pfam; PF00254; FKBP; 1 | | | | | | | | |
| DR | PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG. | | | | | | | | |
| DR | PROSITE; PS00454; FKBP_PPIASE_2; 1 | | | | | | | | |
| DR | PROSITE; PS00059; FKBP_PPIASE_3; 1 | | | | | | | | |
| KW | Isomerase; Rotamase; Nuclear protein; DNA-binding; phosphorylation. | | | | | | | | |
| FT | DOMAIN 91 112 | | | | | | | | |
| FT | DOMAIN 120 145 | | | | | | | | |
| FT | DOMAIN 152 216 | | | | | | | | |
| FT | DOMAIN 219 302 | | | | | | | | |
| FT | DOMAIN 324 412 | | | | | | | | |
| SQ | SEQUENCE 412 AA; 45810 MW; F2A69159AEF4FE22 CRC64; | | | | | | | | |
| Alignment Scores: | | | | | | | | | |
| Pred. No.: | 0.896 | Length: | 412 | | | | | | |
| Score: | 82.50 | Matches: | 32 | | | | | | |
| Percent Similarity: | 41.74% | Conservative: | 16 | | | | | | |
| Best Local Similarity: | 27.83% | Mismatches: | 40 | | | | | | |
| Query Match: | 12.24% | Indels: | 27 | | | | | | |
| DB: | 1 | Gaps: | 5 | | | | | | |
| US-09-936-737A-1 (1-375) x FKBP4_SPOFR (1-412) | | | | | | | | | |
| QY | 43 CTGATGCTCAACTACTCTTCACAGAGACGCTGAGATGTTGGACGCTTTACGGCAGAACGA | 102 | | | | | | | |
| DB | 116 LEVALPRLPRLALYLSASNLYSALRYLSLEUGLNASALALALALALALALALALAL | 135 | | | | | | | |
| QY | 103 AAATATACAGACTTCATTAATCTTTTAAGAGCTCC | 147 | | | | | | | |
| DB | 136 LYSALALYSPRLSPYLSYLSALAGLYLSASNSERIALPRLALALAGLUSERASPSSER | 155 | | | | | | | |
| QY | 148 GACCAATGCAAAAAAACAATGTTCAAGACGAGTACTGCTACATGCTTTTGAAGACACG | 207 | | | | | | | |
| DB | 156 ASPSPASP | 162 | | | | | | | |
| QY | 208 GTCAACAGAGATGTTACTACTACATATGTCGTTGATGCGAAGGTTAGACCAAGAAAATTT | 267 | | | | | | | |
| DB | 163 LEUGLNLSP | 176 | | | | | | | |
| QY | 268 GTTGTCACAGCAAACTTCACAGAAAATTAATTTACAGACTGCGAGCGTAAAGATGACAGT | 322 | | | | | | | |
| DB | 177 ---ASP | 193 | | | | | | | |
| QY | 328 AATGCGCAGGTACAGGTGACGACGACTGATGAAAGTTGATGAGAT | 372 | | | | | | | |
| DB | 194 GLUGLNASPASPASPASPGLUGLNASPGLUGLNASPGLUGLNASPASPASP | 208 | | | | | | | |
| RESULT 3 | | | | | | | | | |
| ID | YG46_YEAST | STANDARD: | PRT; | 507 AA. | | | | | |
| AC | P53301; | | | | | | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | | | | | | |
| DT | 01-OCT-1996 (Rel. 34, Last annotation update) | | | | | | | | |
| DE | 01-NOV-1997 (Rel. 35, Last annotation update) | | | | | | | | |
| GN | Hypothetical 52.8 kDa protein in BURL-HIP1 intergenic region. | | | | | | | | |
| OS | YGR189C OR G7553. | | | | | | | | |
| OC | Saccharomyces cerevisiae (Baker's yeast). | | | | | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | | | | | |
| OX | NCBI_TaxID=4932; | | | | | | | | |
| RA | SEQUENCE FROM N.A. | | | | | | | | |
| RC | STRAIN=5288C; | | | | | | | | |

```

CC -I- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/cc
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72874; CAA97215.1; -.
DR EMBL; X99074; CAA67525.1; -.
DR HSSP; P23904; IABK.
DR SGD; S0003421; CRH1.
DR InterPro; IPR000757; Glyco_hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
KW Hypothetical protein.
FT DOMAIN 63 66 POLY-SER.
FT DOMAIN 301 310 POLY-SER.
FT DOMAIN 345 357 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 467 470 POLY-SER.
SQ SEQUENCE 507 AA; 52757 MW; 7DB61F57AE942C CRC64;

Alignment Scores:
Pred. No.: 1.85 Length: 507
Score: 79.50 Matches: 42
Percent Similarity: 43.18% Conservative: 15
Best Local Similarity: 31.82% Mismatches: 50
Query Match: 12.21% Indels: 25
DB: 1 Gaps: 4

US-09-936-737A-1 (1-375) x YG46_YEAST (1-507)
QY 371 TCTTCATCAACT-----TCATCTGCATCGTCACGACGTACGCGCATTAACCTGCATCT 318
   |||||
Dd 308 SerSerSerThrValSerSerSerValSerSerThrValSerSerSerValSerSerThr 327
   ::|||
QY 317 TTACCTGCGCAGCTCTGTCAATAATTTCGCCGGAACATTTTCGCGCAACAAATTTTCT 258
   ::|||
Dd 328 ValSerSerSerValSerSerThrValSerSerSerValSerSerSerThrValSerSer 347
   ::|||
QY 257 TGGTAACTCCTTCACACCAACAGCATTTGATGAATTCCTCTTGACCGGTCTTCA 198
   |||||
Dd 348 SerSerValSerSerSerSerSerThrSerProSerSerSerThrAlaThr---SerSer 366
   |||||
QY 197 AAACGATGACGACGACTCCGCTTG----- 171
   |||||
Dd 367 LysThrLeuAlaSerSerSerValThrThrSerSerSerIleSerSerPheGluYsGln 386
   ::|||
QY 170 -----AAACATGTTTTTCGATTCGCAAGACAGACGAGCATCTTAA 126
   |||||
Dd 387 SerSerSerSerSerLysThrValAlaSerSerSerThrSerGluSerIleIleSer 406
   |||
QY 125 GAT-----TTATCGAAGCTGTATATTTCGTGCGGTAAACGTC 84
   ::|||
Dd 407 SerThrLysThrProAlaThrValSerSerThrThrArgSerThrValAlaProThrThr 426
   ::|||
QY 83 CAACATCTTCACGTTCTTCTGAAGAAGTGTGAG 48
   |||||
Dd 427 GlnGlnSerSerValSerSerAspSerProValGln 438
   |||||

RESULT 4
BMR2_MOUSE STANDARD; PRT; 1038 AA.
AC 035607;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bone morphogenetic protein receptor type II precursor (EC 2.7.1.37)
DE (BMP type II receptor) (BMPR-II) (BRK-3).
GN BMPR2.
OS Mus musculus(Mouse).

```

| | |
|---------------------------|--|
| CC | Enkaryota; Metazoa; Chordata; Crniata; Vertebrata; Euteleostomi; |
| CC | Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus. |
| CC | NCBI_TaxID=10090; |
| CC | SEQUENCE FROM N.A. |
| CC | MEDLINE=97350808; PubMed=9207184; |
| CC | Beppu H., Mitowa O., Miyazono K., Kawabata M.; |
| CC | "cDNA cloning and genomic organization of the mouse BMP type II |
| CC | receptor "; |
| CC | Biochem. Biophys. Res. Commun. 235:499-504(1997). |
| CC | [2] |
| CC | SEQUENCE FROM N.A. |
| CC | Whitaker G.B., Koenig B.B., Ting J., Tiesman J.P., Limberg A.L., |
| CC | Grant R.A., Begley K.B., Rosenbaum J.S.; |
| CC | "Identification of BMP receptor complexes with differential signaling |
| CC | properties and ligand binding profiles."; |
| CC | Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases. |
| CC | -1- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4. |
| CC | BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS |
| CC | FOR BMPs. |
| CC | -1- CATALYTIC ACTIVITY: ATP + a protein - ADP + a phosphoprotein. |
| CC | -1- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTORS. |
| CC | -1- SUBCELLULAR LOCATION: Type I membrane protein. |
| CC | -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. |
| CC | ----- |
| CC | TPR RECEPTOR SUBFAMILY. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation |
| CC | at the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (See http://www.isb-stb.ch/announce/) |
| CC | or send an email to license@isb-stb.ch) |
| CC | ----- |
| DR | EMBL; AF003942; AAB63042.1; .. |
| DR | EMBL; U78048; AAB87638.1; .. |
| DR | MGD; MGI:1095407; Bmp2r2. |
| DR | InterPro: IPR000472; Activin_rec. |
| DR | InterPro: IPR000719; Euk.pkinase. |
| DR | InterPro: IPR002290; Ser_thr_pkinase. |
| DR | Pfam: PF00069; pkinase; 1. |
| DR | Pfam: PF01064; Activin_rec; 1. |
| DR | ProDom: PDO00001; Euk_pkinase; 1. |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. |
| DR | PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG. |
| DR | PROSITE; PS50011; PROTEIN_KINASE_DOM; 1. |
| KM | Receptor; Transfease; Serine/threonine-protein kinase; ATP-binding; |
| KM | Transmembrane; Glycoprotein; Signal. |
| KM | SIGNAL |
| FT | CHAIN |
| FT | 27 1038 |
| FT | 1 26 |
| FT | POTENTIAL. |
| FT | BONE MORPHOGENETIC PROTEIN RECEPTOR TYPE |
| FT | II. |
| FT | DOMAIN |
| FT | 27 150 |
| FT | 151 171 |
| FT | POTENTIAL. |
| FT | EXTRACELLULAR (POTENTIAL). |
| FT | DOMAIN |
| FT | 172 1038 |
| FT | 203 504 |
| FT | CYTOPLASMIC (POTENTIAL). |
| FT | DOMAIN |
| FT | 209 217 |
| FT | NP_BIND |
| FT | 230 230 |
| FT | ACT_SITE |
| FT | 333 333 |
| FT | BY SIMILARITY. |
| FT | DOMAIN |
| FT | 191 194 |
| FT | POLY-ALA. |
| FT | 547 550 |
| FT | POLY-SER. |
| FT | DOMAIN |
| FT | 610 618 |
| FT | POLY-THR. |
| FT | DOMAIN |
| FT | 901 908 |
| FT | POLY-ASN. |
| FT | CARBOHYD |
| FT | 55 55 |
| FT | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD |
| FT | 110 110 |
| FT | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD |
| FT | 126 126 |
| FT | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| SO | SEQUENCE |
| SO | 1038 AA; 115019 MW; 4106945DC63250E1 CRC64; |
| Alignment Scores: | |
| Pred. No.: | |
| Score: | |
| Best Similarity: | |
| Percent Local Similarity: | |
| Query Match: | |
| Length: | |
| Matches: | |
| Conservative: | |
| Mismatch: | |
| Indels: | |
| 1.94 | |
| 79.50 | |
| 40.46% | |
| 20.61% | |
| 11.80% | |
| 1038 | |
| 27 | |
| 36 | |
| 35 | |
| 43 | |


```

Taich A.;
Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation in
the European Bioinformatics Institute. There are no restrictions on its use
by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce
or send an email to license@isb-sib.ch).
-----
CC      EMBL: U21308; AAB93316.1;
DR      Wormpep; ZK1290.10; CE02080.
DR      InterPro; IPR003582; SHKT.
DR      SMART; SM00254; SHKT; 1.
KW      Hypothetical protein; Signal.
FT      SIGNAL              1      21      POTENTIAL.
FT      CHAIN              1      360      HYPOTHETICAL PROTEIN ZK1290.10.
SQ      SEQUENCE          360 AA; 39669 MW; 18751B6BC062DAF7 CRC64;
-----
Alignment Scores:
Pred. No.:          5 3
Score:              75 00
Percent Similarity: 46.38%
Best Local Similarity: 33.33%
Query Match:        11.13%
Gaps:                1
Indels:              14
Matches:             23
Conservative:        9
Mismatch:            23
Indels:              14
Gaps:                3
-----
US-09-936-737A-1 (1-375) x YOFA.CABEL (1-360)
QY      199 GAGGACAGCGTCACACAGGAAGTGTACTACATGCTGTGATGGTGAGAGTTAGACCAA 258
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      224 GUAAPSTHVAIGLYGLUIGLU-----ASNVALPHEASPHENASPLYSLEUPHEASP 240
QY      259 GAAATATTTGTTGTC-----GACGAAACTTCACGGAATAATTATTG 300
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      241 LysLYLLEtYrILearGAsnAspLYSerThrThrGluAsnThrThrGluGlnSerThr 260
QY      301 ACAGACTCGGAGCGAAGAAAGATGACGTAAT-----GGCGCAGGTACAGT 345
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      261 ThrGluLysThrGlyThrLysLysGlyThrIleGluValSerValGluLeuGlyGly 280
QY      346 GACGAGTCAGATGAAGTTGATGAAGAT 372
DB      281 AspaSpaspGluGluAsnAspaSpasp 289
-----
RESULT 10
MOK1_SCHPO STANDARD: PRT; 2410 AA.
ID MOK1_SCHPO
AC Q9USK8; Q9URT5; Q9URK2;
DE 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase mok1 (EC 2.4.1.183).
DE MOK1 OR AGS1 OR SPC338.01C OR SPC317A7.01.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RS SEQUENCE FROM N.A.
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;
RA "Fission yeast alpha-1,3-glucan synthase MOK1 localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RS SEQUENCE FROM N.A.
RA STRAIN=972.
RA MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

```



```

Oy      259 GAAAAATT-----GTTGTGCAGGAACCTTCACGGAAAT-----TATTGGACA   303
          |||               |||:::||||:|||||||
Db      706 TyAlA1PhelYalLeuTYrValProglInGuAsPheIleGlAasnaSProlleIeThr   725
Oy      304 GA CTGGCAGGCGTAAGATGCAGGTATGCCGCGACGTACAGCGTAGCTACAGT   357
          ::|               ::|:::||||:|||||||
Db      726 SerLeuthrProglInHisAspaLaarGaValJa1aLaseRcLYaSnGUllleASP   743

RESULT 11
ID PA21_CANFA STANDARD: PRT: 146 AA.
AC P06596;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (phosphatidylcholine 2-acylhydrolase) (group IB phospholipase A2).
DE PLA3B.
OS Canis familiaris (Dog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OC NCBI_TaxID=9615;
OX ||]
RN SEQUENCE FROM N.A.
RP MEDLINE=6223862; PubMed=3754861; Tsuruta Y., Fujii Y., Shin M., Ohara O., Temaki M., Nakamura E., Terakita H., Okamoto M.; "Dog and rat pancreatic phospholipases A2: complete amino acid sequences deduced from complementary DNAs."; J. Biochem. 99:733-739(1986).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=90306027; PubMed=2142076; Kerlebec B., Laforge K.S., Vasilopoulos P., Pulgserver A., Scheele G.A.; "Isolation and sequence of the canine pancreatic phospholipase A2 gene."; Eur. J. Biochem. 190:299-304(1990).
RL [3]
RN SEQUENCE FROM N.A.
RP MEDLINE=87175472; PubMed=3562437; Keefelec B., Latorge K.S., Pulgserver A., Scheele G.A.; "Primary structures of canine pancreatic lipase and phospholipase A2 messenger RNAs."; Pancreas 1:430-437(1986).
RL -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-acyl groups in 3-sn-phosphoglycerides.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-acylglycerophosphocholine + a fatty acid anion.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).
CC -----
CC DR EMBL; D00035; BAAN0023.1; -
CC DR EMBL; M35301; AAA30883.1; -
CC DR PIR; A24392; PSD6.
CC DR PIR; S11316; S11316.
CC DR HSSP; P00592; 4P2P.
CC DR InterPro; IPR001211; PhospholipaseA2.
CC DR Pfam; PF00068; phoslip_1.
CC DR ProDom; PD000303; PhospholipaseA2_1.
CC DR SMART; SM00085; PA2c; 1.
CC DR PROSITE; PS00118; PA2_HIS_1.
CC DR PROSITE; PS00119; PA2_ASP_1.
CC DR Hydrolase; Lipid degradation; Calcium; Pancreas; Signal.
KW

```

| | | | |
|------------------------|--------|-------------|-----|
| Score: | 5.63 | Length: | 146 |
| Percent Similarity: | 74.50 | Matches: | 29 |
| Best Local Similarity: | 38.338 | Conserves: | 19 |
| Query Match: | 24.17% | Mismatches: | 37 |
| DB: | 11.058 | Indels: | 35 |
| | 1 | Gaps: | 5 |

03-03-936-737A-1 (1-375) x PA21-CANFA (1-146)

1 MetLyspheiLeuValTCTCTCTCAGAGAA-----6

70 -----CGTGAAGATTGTTGGACCTCCTT-----ThrValAlaAlaAlaIagIuglycyl

19 IleserProargalavalTIPGlnPheargscvtr-----1

112 GACCTCGATAAATCTTTAG-----gasmucellielyscysthrileprogluser 3

```

39 AspProLeuLysAspTyrAsnAspTyrC1ycystTyrCysG1vLeuG1vct1ucacg1  :::
                                     -----AAGTCCTCT 1

```

142 GATCTTGACGGAATGCAAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTTGGAA 50
:::||||| ||| |||
50 - - - - - 20

[illegible]

75 ----- : : : : :
TACTACATGTCGTGATGCTGAAGAGTTAGACCAGA 26

262 AAATTTGTTTCGACGAAACCTTCACCGGTTT-----Serg1uAla1ylsLysleuAspSerCys 83

84 LysPheLeuLeuAspAsnProTyrThrLVSITetGscGttaaGAT 32

21-PTG

PA21_PIG
P00592;
STANDARD;
PRT; 146 AA.

| | |
|-------------|-----------------------------------|
| 21-JUL-1986 | (Rel. 01, Created) |
| 01-JAN-1988 | (Rel. 05, last sequence modified) |
| 15-JUN-2003 | (Rel. 05, last sequence modified) |

Phospholipase A2, major isoenzyme precursor (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase)

PLA2G1B. (Group IB phospholipase A2).
Sus scrofa (Pig).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Sauria

```

sequence from n1
[1]
next_taxid=9823;
Suidae; Sus.

```

TISSUE=Pancreas;
MEDLINE=87231066; PubMed=3300720

de Geus P., van den Bergh C.J., Kuipers O., Verheij H.M.,
Hoekstra W.P.M., de Haas G.H.

Expression of porcine pancreatic phospholipase A₂. Generation of

Journal of Management Inquiry 22(4) 403-420

100

RT active enzyme by sequence-specific cleavage of a hybrid protein from
RT Escherichia coli." ;
RL Nucleic Acids Res. 15:3743-3759(1987).
RL [2]
RP SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RX MEDLINE=87132925; PubMed=3028739;
RA Sellmeyer J.J., Randall T.L., Yamanaka M., Johnson L.R.;
RT "Pancreatic phospholipase A2: isolation of the human gene and cDNAs
RL from porcine pancreas and human lung." ;
RL DNA 5:519-527(1996).
RL [3]
RP SEQUENCE OF 16-146.
RX MEDLINE=71014235; PubMed=5528841;
RA de Haas G.H., Slotboom A.J., Bonsel P.P.M., van Deenen L.L.M.,
RA Maroux S., Pulgarver A., Desnuelle P.;
RT "Studies on phospholipase A and its zymogen from porcine pancreas. I.
RL The complete amino acid sequence." ;
RL Biochim. Biophys. Acta 221:31-53(1970).
RL [4]
RP REVISIONS.
RX MEDLINE=7722092; PubMed=684127;
RA Puljk W.C., Verheij H.M., de Haas G.H.;
RT "The primary structure of phospholipase A2 from porcine pancreas. A
RL biochim. Biophys. Acta 492:254-259(1977).
RL [5]
RP DISULFIDE BONDS.
RX MEDLINE=71014236; PubMed=4919729;
RA de Haas G.H., Slotboom A.J., Bonsel P.P.M., Nieuwenhuizen W.,
RA van Deenen L.L.M., Maroux S., Dionia V., Desnuelle P.;
RT "Studies on phospholipase A and its zymogen from porcine pancreas.
RL I. The assignment of the position of the six disulfide bridges." ;
RL Biochim. Biophys. Acta 221:54-61(1970).
RL [6]
RP ACTIVATION OF IYS-78.
RC TISSUE=Pancreas;
RX MEDLINE=89255488; PubMed=2498336;
RA Tomasselli A.G., Hui J., Fisher J., Zuercher-Neely H., Reardon H.M.,
RA Oriakti E., Keady F.J., Helmrison R.L.;
RT "Dimerization and activation of porcine pancreatic phospholipase A2
RL via substrate level acylation of porcine pancreatic phospholipase A2
RL J. Biol. Chem. 264:10041-10047(1989).
RL [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND ACTIVE AND BINDING SITES.
RX MEDLINE=83368704; PubMed=6876174;
RA Dijkstra B.W., Rensema R., Kalk K.H., Hol W.G.J., Drenth J.;
RT "Structure of porcine pancreatic phospholipase A2 at 2.6-A resolution
RL and comparison with bovine phospholipase A2." ;
RL J. Mol. Biol. 168:163-179(1983).
RL [8]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=91073405; PubMed=2254938;
RA Thunnissen M.M.G.M., Kalk K.H., Drenth J., Dijkstra B.W.;
RT "Structure of an engineered porcine phospholipase A2 with enhanced
RL activity at 2.1-A resolution. Comparison with the wild-type porcine
RL and Crotalus atrox phospholipase A2." ;
RL J. Mol. Biol. 216:425-439(1999).
RL [9]
RP STRUCTURE BY NMR.
RX MEDLINE=91175768; PubMed=2007145;
RA Dekker N., Peters A.R., Slotboom A.J., Boelens R., Kaptein R.,
RA de Haas G.H.;
RT "Porcine pancreatic phospholipase A2: sequence-specific 1H and 15N
RL NMR assignments and secondary structure." ;
RL Biochemistry 30:3135-3147(1991).
RL [10]
RP STRUCTURE BY NMR.
RX MEDLINE=95393214; PubMed=7664098;
RA van den Berg B., Tessari M., Boelens R., Dijkman R., de Haas G.H.,
RA Kaptein R., Verheij H.M.;
RT "NMR structures of phospholipase A2 reveal conformational changes
RL during interfacial activation." ;

| | | | | |
|----|----------|---------|-----------|-------------------------|
| FT | CA_BIND | 54 | 54 | VIA CARBOXYL OXYGEN. |
| FT | CA_BIND | 71 | 71 | |
| FT | LIPID | 78 | 78 | N-OCTANOATE. |
| FT | HELIX | 24 | 34 | |
| FT | HELIX | 36 | 37 | |
| FT | HELIX | 40 | 44 | |
| FT | TURN | 45 | 46 | |
| FT | STRAND | 47 | 47 | |
| FT | TURN | 48 | 50 | |
| FT | STRAND | 51 | 51 | |
| FT | HELIX | 62 | 78 | |
| FT | TURN | 79 | 79 | |
| FT | HELIX | 81 | 84 | |
| FT | TURN | 85 | 86 | |
| FT | TURN | 90 | 92 | |
| FT | STRAND | 97 | 100 | |
| FT | TURN | 101 | 102 | |
| FT | STRAND | 103 | 106 | |
| FT | TURN | 108 | 109 | |
| FT | HELIX | 112 | 130 | |
| FT | HELIX | 135 | 137 | |
| FT | STRAND | 138 | 138 | |
| FT | TURN | 139 | 139 | |
| FT | HELIX | 142 | 145 | |
| SO | SEQUENCE | 146 AA; | 16279 MM; | DE87674C9476FA36 CRC64; |

| | | |
|------------------------|--------|------------------|
| Alignment Scores: | | |
| Pred. No.: | 5.63 | Length: 146 |
| Score: | 74.50 | Matches: 29 |
| Percent Similarity: | 33.06% | Conservative: 11 |
| Best Local Similarity: | 23.97% | Mismatches: 44 |
| Query Match: | 11.05% | Indels: 37 |
| DB: | 1 | Gaps: 4 |

| | |
|--|--|
| US-09-936-737A-1 (1-375) x PA21_PTIG (1-146) | |
| QY 16 ATTTCTCTTCCTTGGCTGGCAAGCTGTGATCTCAACTACTTCTTCAGAA-----GAA 69 | |
| Db 1 MetLysPheLeuValLeuValLeuLeuThrValGlyAlaIaIaGlnGlnGlyIleSer 20 | |
| QY 70 CGTGAAGATTGTGGACGCTTTTACGCGACAGAA----- 105 | |
| Db 21 SerArgAlaLeuThrProIaPheArgSerMetIleLysCysAlaIleProIySerHisPro 40 | |
| QY 106 -----TATACAGACTTCGATTAATCTTTTAAGAAGTCC 136 | |
| Db 41 LeuMetAspPheAsnAsnTyrGlyCysTyrCysGlyLeuGlyGlySerGlyThrProVal 60 | |
| QY 139 TCGATCTTCGAGAAATGCAAAAAACATGTTTCAAGCGSAGTACGCTACATCGTTT 198 | |
| Db 61 AspGlnLeuAspArgCysCysGlnThr----- 69 | |
| QY 199 GAAGACACGGTCAACACAGAAATGTTACTACAAATGCTGTCAGTGAACAGTTAGACCA 258 | |
| Db 70 -----HisAspAsnCysTyrArg-----AspAlaIaLysAsnLeuAspSer 82 | |
| QY 259 GAAAAATTGTCTGTGACAGAAACTTCACGGAATATATTGACAGACTCGCAGGGTTAA 318 | |
| Db 83 CysLysPheLeuValAspAsnProTyrThrGlnSerTyrSerTyrSerCysSerAsnThr 102 | |
| QY 319 GAT 321 | |
| Db 103 Glu 103 | |

| | |
|--|------------------------|
| RESULT 13 | |
| Z222_HUMAN | |
| ID Z222_HUMAN | STANDARD; PRT; 451 AA. |
| AC O9UK12; | |
| DT 16-OCT-2001 (Rel. 40, Created) | |
| DT 16-OCT-2001 (Rel. 40, Last sequence update) | |
| DT 16-OCT-2001 (rel. 40, Last annotation update) | |
| DE Zinc finger protein 222. | |
| ZN ZNF222. | |


```

OY 178 GAGTACGCTGACAGCTGTTTTGAA-----GACAGCGGCACCAAGGAATGTAC 225
    ||| |||||:::
Db 48 SerTyrLysTYrValAspTyrLysProIleGlyGlyIuIleIleIleLysAspValTyr 67
OY 226 TACATGTCGTTGAGCGGAAGACTTAGACCAAGAAAAATTTGTCGCGAGAAACTTC 285
    ||| ::||| ||||| |||:::
Db 68 TyrArgAspArgAsnGlyGluGluPheLysIleGluGluIleIleIleGluLys----- 85
OY 286 ACGGAAATTTATTTGACAGACTCGAGGAGGTAA 318
    |||::: |||:::
Db 86 -----LeuSerGluThrGluGlyLys 92

```

```

Db      508  SerSerSerAlaSerSerSerSerAlaProSerSerSerAlaSerSerSerAlaSer  527
Oy      251  AACCTTCACCATCACAG  234
          :::|||||  |||:::
Db      528  SerSerSerAlaSerSer  533

```

Search completed : February 25, 2003, 02:26:31
Job time : 16.5 secs

Search completed: February 25, 2003, 02:26:31
Job time : 16.5 secs

| ID | SP96_DICTID | STANDARD: | PRT: | 600 AA. |
|----|--|-----------------------------------|------|---------|
| AC | P14328: | | | |
| DT | 01-JAN-1990 | (Rel. 13, Created) | | |
| DT | 01-JAN-1990 | (Rel. 13, Last sequence update) | | |
| DT | 01-NOV-1997 | (Rel. 35, Last annotation update) | | |
| DE | Spore coat protein SP96. | | | |
| GN | COT4. | | | |
| OS | Dictyostelium discoideum (Slime mold). | | | |
| OC | Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium. | | | |
| OX | NCBI_TaxId=44689; | | | |
| RM | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-AX4; | | | |
| RX | MEDLINE=90067962; PubMed=2587278; | | | |
| RA | Fosnaugh K., Loomis W.F.; | | | |
| RT | "Sequence of the Dictyostelium discoideum spore coat gene SP96."; | | | |
| RL | Nucleic Acids Res. 17:9489-9489(1989). | | | |
| CC | -1 SUBCELLULAR LOCATION: OUTER LAYER OF THE COAT MATRIX AND INTERSPORE MATRIX. | | | |
| CC | -1 PTM: PHOSPHORYLATED AND PHOSPHOSYLATED, MAY BE PHOSPHOGLYCOSYLATED, MAY CONTAIN GLYCAC-ALPHA-1-P-SER RESIDUES. | | | |
| CC | -1 SIMILARITY: CONTAINS 4 PRESPORE MOTIFS. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL, X16491; CAA34508.1; - | | | |
| DR | PIR, S07638; S07638. | | | |
| DR | Glycositedb: P14328; - | | | |
| DR | Dictyodb: DD03007; cotA. | | | |
| DR | InterPro, IPR003645; FOLN. | | | |
| DR | SMART, SM00274; FOLN: 6. | | | |
| KW | Glycoprotein; Phosphorylation; Repeat; Sporulation. | | | |
| FT | REPEAT 185 197 PRESPORE MOTIF 1. | | | |
| FT | REPEAT 221 233 PRESPORE MOTIF 2. | | | |
| FT | REPEAT 298 310 PRESPORE MOTIF 3. | | | |
| FT | REPEAT 395 407 PRESPORE MOTIF 4. | | | |
| SQ | SEQUENCE 600 AA: 59589 MW: 616AEDD02B5F1071 CRC64; | | | |

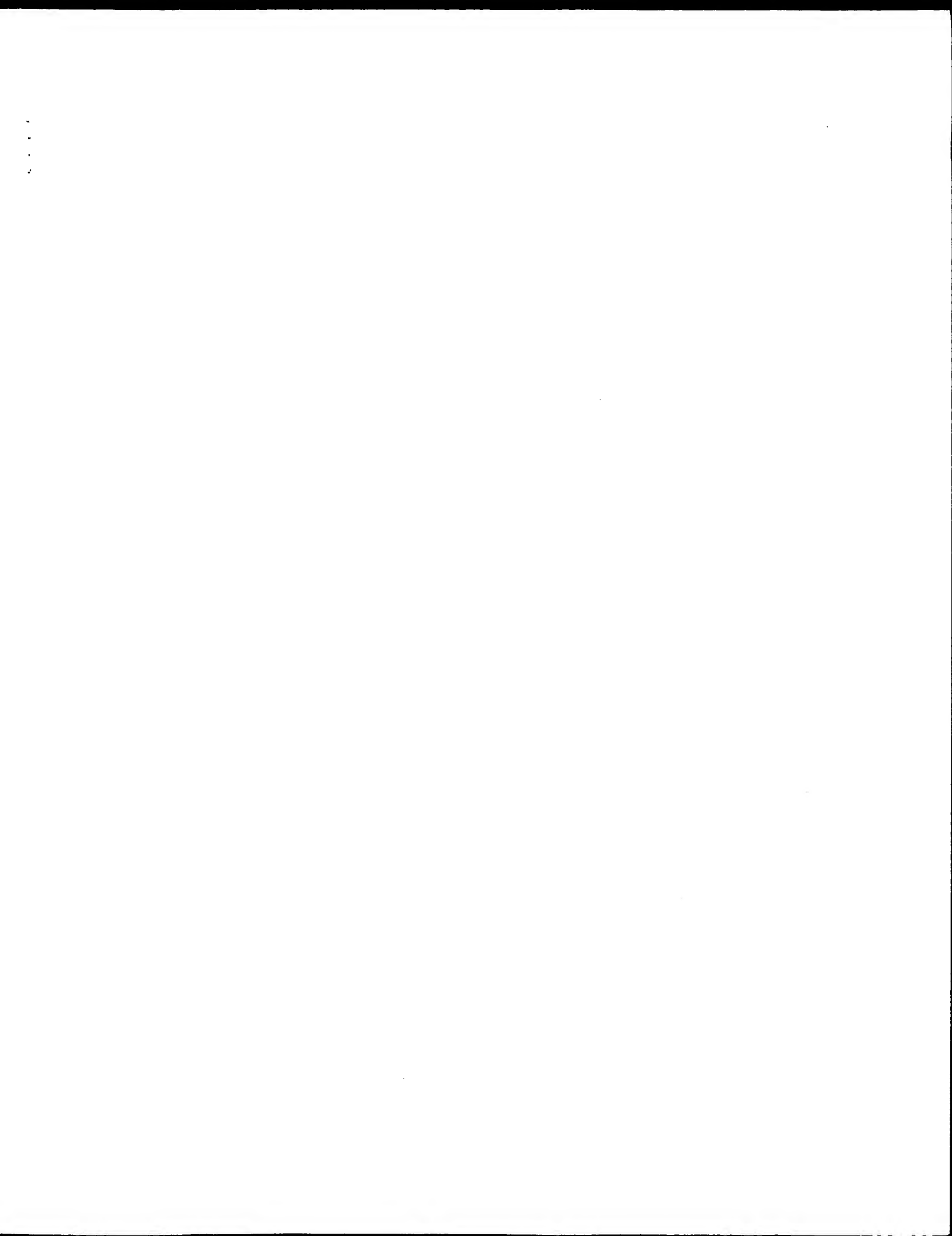
| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 6.95 |
| Score: | 74.00 |
| Percent Similarity: | 58.70% |
| Best Local Similarity: | 45.65% |
| Query Match: | 11.37% |
| OB: | 1 |
| | |
| length: | 600 |
| Matches: | 21 |
| Conservative: | 6 |
| Mismatches: | 19 |
| Indels: | 0 |
| Gaps: | 0 |

US-09-936-737A-1 (1-375) x SP96_DICDI (1-600)

```

0y 311 TCTTATCAACTTCATCTGCATCGTCGACCTTACTCGCGCATTAACGATCTTTTACC 312
      |||||  ||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 488 SerSerSerAlaSerSerSerSerSerProSerSerSerAlaSerSerSerSerAlaPro 507
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0y 311 TCGCAGCTCGTCAAAATATTTTCCGTGAAGTTTCCTCGACACAAATTTTTCTTGCTCT 254

```



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:54:30 ; Search time 44 Seconds
(without alignments)
3512.167 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 674
Sequence: 1 atgaagtattcttgatttc.....atgaagtatgaatgaattaa 375

Scoring table:
BLOSOM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+nrp.model -DEV=xlh
-Q/cq92.1/USPTO.SP001/US09936737/runat.14022003.100557.1398/app-query.fasta.1.519
-DB=SPREMBL.21 -QEXT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=pmo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -ICPU=3
-USRR=US09936737 @CGN.1.1.34 @runat.14022003.100557.1398 -NCPU=6 -ICPU=3
-NO_XLTRY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL.21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriophage:*
17: sp.archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|-------|---------------------|
| 1 | 89.5 | 13.3 | 3933 | 5 097239 plasmodium |

| | | | | | | |
|----|------|------|------|----|--------|--------------------|
| 2 | 84.5 | 12.5 | 164 | 5 | 044835 | 044835 caenorhabdi |
| 3 | 82.5 | 12.2 | 368 | 5 | 09vfh8 | 09vfh8 drosophila |
| 4 | 82.5 | 12.2 | 374 | 5 | 085xc5 | 085xc5 drosophila |
| 5 | 82 | 12.6 | 637 | 10 | 080625 | 080625 arabidopsis |
| 6 | 82 | 12.2 | 907 | 5 | 09xts9 | 09xts9 caenorhabdi |
| 7 | 81 | 12.0 | 388 | 16 | 08r7b3 | 08r7b3 thermoaer |
| 8 | 80.5 | 11.9 | 1303 | 10 | 09sk50 | 09sk50 arabidopsis |
| 9 | 79.5 | 12.2 | 550 | 10 | 094by9 | 094by9 arabidopsi |
| 10 | 79.5 | 11.8 | 1277 | 11 | 035821 | 035821 rattus norv |
| 11 | 79 | 11.7 | 358 | 5 | 08sra3 | 08sra3 encephalit |
| 12 | 78.5 | 11.6 | 116 | 16 | 08xj41 | 08xj41 clostridium |
| 13 | 78.5 | 11.6 | 264 | 5 | 020246 | 020246 caenorhabdi |
| 14 | 78.5 | 11.6 | 509 | 2 | 09x9s1 | 09x9s1 streptococc |
| 15 | 78.5 | 11.6 | 550 | 5 | 09u204 | 09u204 caenorhabdi |
| 16 | 78.5 | 11.6 | 682 | 16 | 09k8n5 | 09k8n5 bacillus ha |
| 17 | 77.5 | 11.5 | 191 | 5 | 020522 | 020522 caenorhabdi |
| 18 | 77 | 11.4 | 257 | 5 | 09n5s7 | 09n5s7 caenorhabdi |
| 19 | 77 | 11.4 | 535 | 16 | 09sp06 | 09sp06 mycoplasma |
| 20 | 77 | 11.4 | 945 | 5 | 003650 | 003650 trypanosoma |
| 21 | 76.5 | 11.4 | 168 | 10 | 09l013 | 09l013 arabidopsis |
| 22 | 76.5 | 11.4 | 343 | 8 | 08w921 | 08w921 demansia at |
| 23 | 76.5 | 11.4 | 1255 | 16 | 09c7e7 | 09c7e7 arabidopsis |
| 24 | 76 | 11.3 | 1278 | 16 | 050667 | 050667 borrelia bu |
| 25 | 76 | 11.3 | 336 | 10 | 09s200 | 09s200 arabidopsi |
| 26 | 75.5 | 11.2 | 509 | 8 | 09tkr9 | 09tkr9 lindenbergt |
| 27 | 75.5 | 11.2 | 703 | 10 | 09lkr9 | 09lkr9 arabidopsi |
| 28 | 75.5 | 11.2 | 719 | 16 | 099263 | 099263 streptococc |
| 29 | 75.5 | 11.2 | 210 | 3 | 09usb7 | 09usb7 schizosacch |
| 30 | 75 | 11.1 | 324 | 2 | 09s0m3 | 09s0m3 staphylococ |
| 31 | 75 | 11.1 | 344 | 3 | 09t109 | 09t109 schizosacch |
| 32 | 75 | 11.1 | 534 | 3 | 094317 | 094317 schizosacch |
| 33 | 75 | 11.1 | 782 | 10 | 09lkr5 | 09lkr5 arabidopsi |
| 34 | 75 | 11.1 | 2410 | 3 | 09r852 | 09r852 schizosacch |
| 35 | 75 | 11.1 | 3844 | 5 | 094648 | 094648 plasmodium |
| 36 | 74.5 | 11.1 | 146 | 10 | 08s554 | 08s554 rosa hybrid |
| 37 | 74.5 | 11.1 | 156 | 11 | 09c049 | 09c049 mus muscul |
| 38 | 74.5 | 11.1 | 328 | 3 | 012140 | 012140 saccharomyc |
| 39 | 74.5 | 11.1 | 451 | 4 | 09p105 | 09p105 homo sapien |
| 40 | 74.5 | 11.1 | 469 | 2 | 08vwm8 | 08vwm8 listeria se |
| 41 | 74.5 | 11.1 | 606 | 10 | 09f1w8 | 09f1w8 arabidopsi |
| 42 | 74.5 | 11.1 | 643 | 5 | 09xv53 | 09xv53 caenorhabdi |
| 43 | 74.5 | 11.0 | 248 | 5 | 09vr79 | 09vr79 drosophila |
| 44 | 74 | 11.0 | 292 | 12 | 065239 | 065239 african swi |
| 45 | 74 | 11.0 | | | | |

ALIGNMENTS

RESULT 1
ID 097239 PRELIMINARY; PRT; 3933 AA.

AC 097239; 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DI 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 467.9 kDa protein.
GN PFC0245C, MAL3P2.18.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36529;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S.C., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moulie S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.,
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum .";

RL Nature 400:533-538(1999).
 DR EMBL; AL034558; CAB39005.1;
 DR InterPro; IPR002048; EF-hand.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN.1.
 KW Hypothetical protein.
 SQ SEQUENCE 3933 AA; 467876 MW; 5144A4604EE36933 CRC64;

Alignment Scores:

Pred. No.: 0.56 Length: 3933
 Score: 89.50 Matches: 36
 Percent Similarity: 45.52% Conservative: 30
 Best Local Similarity: 24.83% Mismatches: 40
 Query Match: 13.28% Indels: 40
 Gaps: 8

US-09-936-737a-1 (1-375) x 097239 (1-3933)

QY 56 CTTCTCGAAGAACCGAAGATTGTT-----CGACGTTTACGCGA 97
 Db 1086 ILeuLunsnasnlnslnsLeuTyrrheasplysglTyraNglYasnTySerLys 1105
 QY 98 ACAGAAATATACAGACTTCGATAATCTT-----TTAGAGAT 136
 Db 1106 ILeuLunsnasnlnslnsLeuMetLysTyrrheasplysglN 1125
 QY 137 CCTCTGATCTTCGACGAATCAAAAAACATGT-----TTCAAGACGAGTACTGC 186
 Db 1126 AsnLeuLleasplalIlelleLysasnHsIleuTyrrheasplysglN 1145
 QY 187 TACATCGTTT-----GAAGACAGGTCACAC 216
 Db 1146 TyrllellelphleuPhetyrAsnTyrrHsleuLleuLysglLysLleuAsnLys 1165
 QY 217 GAATGTACTAC-----AATGTCGTGATGATGAGAGTAGAC--CAAGAAAA 264
 Db 1166 SerCysPhetyrPhelLysAsnIleleuLsnasnLysThrCysglLysAsnLys 1185
 QY 265 TTT-----GTTGTGACGAAACTTCACGCAAAATAT 297
 Db 1186 PhetyrPhetyrCysPhelLysPhelLysHsIlelleLysglLysAsnLysSerLys 1205
 QY 298 TTGACAGACTCGAGGCGTAAGATGACAGTAGTACGCGACAGTACAGTACGATCAGAT 357
 Db 1206 LeuLysasnTyrrLysIleLysglLysAlaLysAspThrAspAspAspAspAsp 1225
 QY 358 GAAGTTGAGAGAT 372
 Db 1226 AspAspAspAsp 1230

RESULT 2

044835

AC 044835 PRELIMINARY; PRT; 164 AA.

DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE F54D12.1 protein.
 GN F54D12.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsouh R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Garston M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten N., Lalster P., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Tillerly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Magi L., Goela D.;
 RT "The sequence of C. elegans cosmid F54D12.";
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF040647; AAB94992.1;
 SQ SEQUENCE 164 AA; 19130 MW; D3C070A63C493716 CRC64;

Alignment Scores:

Pred. No.: 2.05 Length: 164
 Score: 84.50 Matches: 23
 Percent Similarity: 45.35% Conservative: 16
 Best Local Similarity: 26.74% Mismatches: 34
 Query Match: 12.54% Indels: 13
 Gaps: 3

US-09-936-737a-1 (1-375) x 044835 (1-164)

QY 10 TTCTGTGATTCCTCTGCTTGCCTCGAAGCTGTGATCTGACATCTTCTTGAGAGAA 69
 Db 5 PhelLleuLecystyrrValleuValasnserlleMetValasglYalalProglu 24
 QY 70 CGTGAAGATGTTGACGCTTTACGCGAAGCAAGAAATATACAGCTTCGATTAATCTTT 129
 Db 25 Ileser-----ThrseryserThrgluile 33
 QY 130 AAGATCTCTGATCTGACGATGCAAGCAAAAAATGTTTCAAGACGAGTCTGCTAC 189
 Db 34 IleasnserLeuThrrpAsplucysVallysglnLysleuSerThrgluileCysIle 53
 QY 190 ATCGTTTGAAGACGCGTCAACAGCAAGATGTTAC--TACATGTCGTTGATGTGAA 246
 Db 54 MetAlaTyrserserLysasnLys--lleCystyrrLeuTyrrAlaValGlyaspallile 72
 QY 247 GAGTTAGACCAAGAAAA 264
 Db 73 GluValAlarghHsaspGln 78

RESULT 3

09VFH8

AC 09VFH8 PRELIMINARY; PRT; 368 AA.

DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CG14853 protein.
 GN CG14853.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Manatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 April J.F., Agdayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palzozo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spleer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zhang X.H., Myers E.W., Rubin G.M., Venter J.C., Zhu S., Zhu X., Smith H.O.,
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 DR EMBL: AE003705; AAF5078.1;
 DR FlyBase: FBgn0058246; CG14853.
 SO SEQUENCE 368 AA; 38008 MW; 298927508C12DBFC CRC64;

Alignment Scores:

Pred. No.: 3 25 Length: 368
 Score: 82.50 Matches: 34
 Percent Similarity: 42.14% Conservative: 25
 Best Local Similarity: 24.29% Mismatches: 42
 Query Match: 12.24% Indels: 39
 DB: 5 Gaps: 7

US-09-936-737A-1 (1-375) x Q9VFNH (1-368)

QY 34 GCAAGCTTGGTATCTCACTACTCTTTCAGAGAAGCTGAAGATTGTTGACGTTTAC 93
 DB 71 AAlserGlyLeuProThrSerAlaSerSerGluAspLeuSerGlnSerLeuSer----- 88
 QY 94 GCGAAGCAAGAAATATACAGACTTCGATTAATCT----- 126
 DB 89 -----GluTyrThrAspAlaAspGluSerValSerAlaProThrGluPheLeuAla 105
 QY 127 ---TTTAAGAGCTCTCTGATCTTTCAGCAATGCAAAAACATGTTTCAAGCGAGTAC 183
 DB 106 GluPheLeuSerAlaValMetLeuLysAspTyrLysAla-----LeuLysTyr 122
 QY 184 TGCTACATCGTTTTCGAA-----GACACGGTCAACAGAAAGTTTACTACAAATGTC 234
 DB 123 CysLysLeuLeuLeuGlnTyrGluProAspAsnAlaThrAlaLysGluPheTyrProLeu 142
 QY 235 GTTGATGTGGAAGAGTTTACCAAGAAAATTGTTGTC-----GAGGAA 279
 DB 143 Ile-----LeuAspLysLeuArgAlaValAlaThrSerSerAspSerAspGlu 158
 QY 280 AACTTCACGGAAGAAAT-----TATTGACAGACTGCGAG 312
 DB 159 AsnTyrAsnLysSerSerSerProAspLeuAlaLeuAspLeuHisAlaSerAspValGlu 178
 QY 313 GGTAAAGATGACGTAATGCGGACGATACAGTGACGATGAGTGAAGTGAAGAT 372
 DB 313 GGTAAAGATGACGTAATGCGGACGATACAGTGACGATGAGTGAAGTGAAGAT 372

DB 179 AlaAspValAspGluAspGluAlaGluAspAlaAspGluAspGluAspAlaAsp 198
 RESULT 4
 ID 08SYG5 PRELIMINARY; PRT; 374 AA.
 DB 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN RE63320P.
 GN C14853.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brockstein P., Hong L., Agdayani A., Carlson J.,
 Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
 Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 Celinker S.,
 RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY015172; ALU49194.1;
 SO SEQUENCE 374 AA; 38729 MW; 914936D5E40AF93A CRC64;

Alignment Scores:

Pred. No.: 3 24 Length: 374
 Score: 82.50 Matches: 34
 Percent Similarity: 42.14% Conservative: 25
 Best Local Similarity: 24.29% Mismatches: 42
 Query Match: 12.24% Indels: 39
 DB: 5 Gaps: 7

US-09-936-737A-1 (1-375) x Q8SYG5 (1-374)

QY 34 GCAAGCTTGGTATCTCACTACTCTTTCAGAGAAGCTGAAGATTGTTGACGTTTAC 93
 DB 77 AAlserGlyLeuProThrSerAlaSerSerGluAspLeuSerGlnSerLeuSer----- 94
 QY 94 GCGAAGCAAGAAATATACAGACTTCGATTAATCT----- 126
 DB 95 -----GluTyrThrAspAlaAspGluSerValSerAlaProThrGluPheLeuAla 111
 QY 127 ---TTTAAGAGCTCTCTGATCTTTCAGCAATGCAAAAACATGTTTCAAGCGAGTAC 183
 DB 112 GluPheLeuSerAlaValMetLeuLysAspTyrLysAla-----LeuLysTyr 128
 QY 184 TGCTACATCGTTTTCGAA-----GACACGGTCAACAGAAAGTTTACTACAAATGTC 234
 DB 129 CysLysLeuLeuLeuGlnTyrGluProAspAsnAlaThrAlaLysGluPheTyrProLeu 148
 QY 235 GTTGATGTGGAAGAGTTTACCAAGAAAATTGTTGTC-----GAGGAA 279
 DB 149 Ile-----LeuAspLysLeuArgAlaValAlaThrSerSerAspSerAspGlu 164
 QY 280 AACTTCACGGAAGAAAT-----TATTGACAGACTGCGAG 312
 DB 165 AsnTyrAsnLysSerSerSerProAspLeuAlaLeuAspLeuHisAlaSerAspValGlu 184
 QY 313 GGTAAAGATGACGTAATGCGGACGATACAGTGACGATGAGTGAAGTGAAGAT 372
 DB 185 AlaAspValAspGluAspGluAlaGluAspAlaAspGluAspGluAspAlaAsp 204
 RESULT 5
 ID 080625 PRELIMINARY; PRT; 637 AA.
 AC 080625;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

| | | | |
|--|--|---------------|----------------------------|
| Dt | 01-NOV-1999 (TREMBLrel. | 12, | (Created) |
| Pf | 01-NOV-1999 (TRENBLrel. | 12, | Last sequence update) |
| Dt | 01-MAR-2002 (TREMBLrel. | 20, | Last annotation update) |
| De | y68H2.2 protein. | | |
| Gn | y68H2.2. | | |
| Oc | Ctenophora elgans. | | |
| Oc | Eukaryota; Metazoa; Nematoda; | | Rhabditidae; Rhabditoidea; |
| Ox | NCB1_Taxid=6239; | | |
| Rn | [1] | | |
| Rp | SEQUENCE FROM N.A. | | |
| Rl | Murray A.A.; | | |
| Rl | Submitted (Aug-1997) to the EMBL/genbank/DBJ databases. | | |
| Rn | [2] | | |
| Rp | SEQUENCE FROM N.A. | | |
| Rx | MEDLINE=9069613; PubMed=9851916; | | |
| Ra | none; | | |
| Rt | "genome sequence of the nematode C.elegans: A platform for | | |
| Rt | investigating biology." | | |
| Rl | Science 282;2012-2018(1998). | | |
| Rl | EMBL; Z98677; CABLI570.1; --. | | |
| Dr | HSSP; Q12784; HRE. | | |
| Dr | IInterPro; IPR000561; EGF-like. | | |
| Dr | IInterPro; IPR001438; EGF_II. | | |
| Dr | IInterPro; IPR001873; Na+channel_ASC. | | |
| Dr | pfam; PF00858; ASC; 1. | | |
| Dr | pfam; PF00008; EGF; 7. | | |
| Dr | PRINTS; PR01078; AMINCHANNEL. | | |
| Dr | PRINTS; PR0010; EGFBLOOD. | | |
| Dr | SMART; SM00181; EGF; 8. | | |
| Dr | PROSITE; PS00022; EGF_L; UNKNOWN_8. | | |
| Dr | PROSITE; PS01186; EGF_2; 4. | | |
| Kw | EGF-like domain; Glycoprotein. | | |
| SQ | SEQUENCE 907 AA; 101840 MW; BBS5FFI3PF8FIID7 CRC64; | | |
| A | Alignment Scores: | | |
| Pred. No. : | 3.57 | Length: | 907 |
| Score: | 82.00 | Matches: | 37 |
| Percent Similarity: | 39.85% | Conservative: | 16 |
| Best Local Similarity: | 27.82% | Mismatches: | 40 |
| Query Match: | 12.1% | Indels: | 40 |
| Db: | 5 | Gaps: | 9 |
| US-09-936-737A-1 (1-375) x OXHS9 (1-907) | | | |
| Oy | ATCTCAACTTCCTTGACAGAACGTGACATTGGACGTTT----- | | 90 |
| : ::::: :::: | | | |
| Db | 206 ILGLYTHrHisGIlnThrLyStyrILysAspAsplySMethrSerAspyLSycSaIngu | | 225 |
| Oy | 91 TACCAGAACAATAATKACACTGCATTAACCTTAG----- | | 132 |
| | | | |
| Db | 226 TyralnglnglysserThrp---asplsrnlleulsTyrlLeuthreCylsgly | | 244 |
| Oy | 133 -----AAGCCCTCATGCTT-----GAAGAA | | 153 |
| : ::::: :::: | | | |
| Db | 245 GlutPrCySmethylserGlunSpmetHeuVallysAspAspValserMetylsaspASP | | 264 |
| Oy | 154 TGCAAAAAACATGT-----TTCAAAGCGAGTCTCCTCACATCGTTTTGAAGACAG | | 207 |
| : ::::: :::: | | | |
| Db | 265 CyslSySLyserCylglAlapSerSerSerGluTrCYcs-----glyLYs | | 279 |
| Oy | 208 GTCAACAAGAAGATTACTNCAATGCTGCTGGTGAAGATTAGACCAAGAAAATTTT | | 267 |
| : ::::: :::: | | | |
| Db | 280 ValmsnerarGcySlle-----ValTYrgLuAlaglyGluMeGrLU----- | | 293 |
| Oy | 268 GTTTCAGCAAAAACCTTACCGAAAATTATTGACAGACTCGAGCGTAAGATGACAGT | | 327 |
| : ::::: :::: | | | |
| Db | 294 ---ThrasprprosaalacacycluaSpasnSetrrhlLeucys--GYAlagluduegly | | 311 |
| Oy | 328 AATGCCGAGGTACAGGTACGAGTACAGTACAGTGAAGTTGAT 366 | | |
| ::: :::: | | | |
| Db | 312 HisClmelyslleaSnrttlpglnsertArpValTLtnrAp 324 | | |

RA kalinka M.D., Barbe V., Peyretallade E., Brothier P., Winocur
RA Prensier G.,

Search completed: February 25, 2003, 02:28:11
job time : 50 secs

| RESULT 15 | | |
|---|-------------------------|------------------|
| ID | PRELIMINARY: | PRT: 550 AA. |
| 09U204 | | |
| AC 09U204: | | |
| DC 01-MAY-2000 (TREMBLrel. 13) | Created | |
| DT 01-MAY-2000 (TREMBLrel. 13) | Last sequence update) | |
| DT 01-JUN-2001 (TREMBLrel. 17) | Last annotation update) | |
| DE Y26D4A.2 protein. | | |
| GN Y26D4A.2. | | |
| OS Caenorhabditis elegans. | | |
| OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; | | |
| OC Rhabditidae; Pelodierinae; Caenorhabditis. | | |
| OC NCBI_Taxid=6239; | | |
| OX [1] | | |
| RN SEQUENCE FROM N.A. | | |
| RA White S.; (FEB-1999) to the EMBL/Genbank/DBJ databases. | | |
| RL Submitted | | |
| RL [2] | | |
| RP SEQUENCE FROM N.A. | | |
| RX MEDLINE=99069613; PubMed=9851916; | | |
| RA none; | | |
| RT "Genome sequence of the nematode C.elegans: A platform for | | |
| RL Science 282:2012-2018(1998). | | |
| DR EMBL: AL110478; CAB54340.1; - | | |
| DR InterPro: IPR001304; Lectin_C. | | |
| DR SMART: SM00034; CLECT: 1. | | |
| SQ SEQUENCE 550 AA; 62195 MW; 79AE220EB127466A CRC64; | | |
| Alignment Scores: | | |
| Pred. No.: | 8.44 | Length: 550 |
| Score: | 78.50 | Matches: 23 |
| Percent Similarity: | 38.89% | Conservative: 12 |
| Best Local Similarity: | 25.56% | Mismatches: 32 |
| Query Match: | 11.65% | Indels: 23 |
| | 5 | Gaps: 3 |

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 00:58:30 ; Search time 1963 seconds
(without alignments)
3093.891 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 375
Sequence: 1 atgaagtattcttgattc.....atgaagtgtgaagaatcaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|-----|--------------------|
| 1 | 51.8 | 13.8 | 879 | AZ550718 ENT50718 |
| 2 | 51.6 | 13.8 | 701 | BM168486 EST571009 |
| 3 | 51.6 | 13.8 | 721 | BM165474 EST567997 |
| 4 | 51.2 | 13.7 | 881 | AZ688514 ENT50227F |
| 5 | 50 | 13.3 | 236 | BM163498 EST566021 |
| 6 | 50 | 13.3 | 272 | BM16089 EST568612 |

| | | | | | | |
|----|------|------|------|----|----------|---------------------|
| 7 | 50 | 13.3 | 274 | 13 | BM166886 | EST569409 |
| 8 | 50 | 13.3 | 276 | 13 | BM163293 | EST565816 |
| 9 | 50 | 13.3 | 425 | 13 | BM162507 | EST565030 |
| 10 | 50 | 13.3 | 453 | 13 | BM167376 | EST566899 |
| 11 | 50 | 13.3 | 459 | 13 | BM171330 | EST573853 |
| 12 | 50 | 13.3 | 580 | 13 | BM161175 | EST563698 |
| 13 | 50 | 13.3 | 648 | 13 | BM164042 | EST565655 |
| 14 | 50 | 13.3 | 651 | 13 | BM166448 | EST568971 |
| 15 | 50 | 13.3 | 667 | 13 | BM161293 | EST563816 |
| 16 | 50 | 13.3 | 678 | 13 | BM170834 | EST573357 |
| 17 | 50 | 13.3 | 702 | 13 | BM169291 | EST571814 |
| 18 | 50 | 13.3 | 719 | 13 | BM602689 | EST501779 |
| 19 | 50 | 13.3 | 739 | 13 | BM160442 | EST562965 |
| 20 | 50 | 13.3 | 753 | 13 | BM167115 | EST569638 |
| 21 | 50 | 13.3 | 762 | 12 | BE846498 | PSB 206 L |
| 22 | 50 | 13.3 | 790 | 13 | BM166415 | EST568938 |
| 23 | 49.2 | 13.1 | 942 | 17 | BM148582 | ENTPR627F |
| 24 | 49 | 13.1 | 816 | 17 | AZ535744 | ENTCQ251R |
| 25 | 49 | 13.1 | 891 | 17 | AZ683582 | ENTPK471R |
| 26 | 49 | 13.1 | 976 | 17 | BM149983 | ENTOD931F |
| 27 | 48.4 | 12.9 | 435 | 13 | BM169710 | EST572233 |
| 28 | 47.4 | 12.6 | 900 | 17 | AZ549980 | ENTD944F |
| 29 | 47 | 12.5 | 843 | 17 | AZ551618 | ENTDV547R |
| 30 | 47 | 12.5 | 877 | 17 | AZ531291 | ENTBQ341R |
| 31 | 47 | 12.5 | 908 | 17 | AZ548467 | ENTEK307R |
| 32 | 47 | 12.5 | 912 | 17 | AZ551092 | ENTFJ227F |
| 33 | 47 | 12.5 | 931 | 17 | BM160272 | ENTOV497R |
| 34 | 46.8 | 12.5 | 724 | 13 | BM161274 | EST563797 |
| 35 | 46.8 | 12.5 | 1006 | 17 | AZ679713 | ENTH567R |
| 36 | 46.6 | 12.4 | 558 | 13 | BM163267 | EST565790 |
| 37 | 46.6 | 12.4 | 890 | 17 | BM146886 | ENTPK487F |
| 38 | 46.6 | 12.4 | 906 | 17 | BM153606 | ENTTS837F |
| 39 | 46.4 | 12.4 | 468 | 12 | BG553005 | dab82e09. |
| 40 | 46.4 | 12.4 | 880 | 17 | AZ668474 | ENTY8887R |
| 41 | 46.2 | 12.3 | 493 | 13 | BM170559 | EST573082 |
| 42 | 46.2 | 12.3 | 1101 | 17 | CNS0181N | AL108773 Drosophila |
| 43 | 45.8 | 12.2 | 849 | 17 | AZ546009 | ENTFW337F |
| 44 | 45.8 | 12.2 | 905 | 17 | AZ550256 | ENTEW587R |
| 45 | 45.4 | 12.1 | 916 | 17 | AZ671886 | ENTLR447R |

ALIGNMENTS

RESULT 1
AZ550718/c 879 bp DNA linear GSS 14-NOV-2000
LOCUS ENT50718
DEFINITION Entamoeba histolytica sheared DNA Entamoeba histolytica
ACCESSION AZ550718
VERSION AZ550718.1 GI:11176019
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 879)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL Unpublished (2000)
COMMENT HMI:IMSS sheared DNA library
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 25
High quality sequence stop: 801.

FEATURES

Location/Qualifiers
1. 879

Source

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"

/clone_1lb="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1; Site: 1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT

193 a 219 c 72 g 395 t

ORIGIN

Query Match

13.8%; Score 51.8; DB 17; Length 879;

Best Local Similarity 46.9%; Pred. No. 0.0004;

Matches 161; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```

OY 29 GCCTGCAAGCTTGTCTCAACTCTCTTCAGAGACGAGATGTTGGACGT 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 768 GACTGTGTTACTGCTTTAATGCTTTGCCAGAGAAATGACCTTGCACACAT 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 89 TTTACCGACAGAAATATACAGACTTCGATTAATCTTTTAAAGTCTCGATCTTG 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 708 TCGATTTGATTTCTACGTAGTATACGTAGAGCTTTTGAAGAAATATATGATG 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 149 ACGATGCAAAAACATGTTTCAAGACGAGTCTGTAATGCTTTTAAACACACG 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 648 AAGATGATGATTTCCAAATTAGACGAGAGAGATGATGAAGAATCAAGATGAAG 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 209 TCAACAGAGAAATGTTACTACATGTCGTTGATGAGAGATTGACCAAGAAATTTG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 588 AAGATGAGAGATGACAGATGAAGATGAAGAAATGAAGATGAAGATGAAGATG 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 269 TTGTGACGAAATCTTCACGAAATTTTGTACAGACTGCGAGGTTAAGATCCAGTA 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 528 AAGAAGATGAAGAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGAG 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 329 ATGCGCAGGTACAGGTGACGATGAGTGAAGTTGATGAAGA 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 468 ATGAAGAAGATGAAGAGATGAAGATGAAGATGAAGA 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2

BM168486

LOCUS

DEFINITION BM168486 701 bp mRNA linear EST 04-DEC-2001

ESR571009 PYBS Plasmodium yoelii yoelii cDNA clone pYCPB36 5' end,

mRNA sequence.

ACCESSION

BM168486

VERSION

BM168486.1 GI:17301718

KEYWORDS

EST.

SOURCE

Plasmodium yoelii yoelii.

ORGANISM

Plasmodium yoelii yoelii.

REFERENCE

1 (bases 1 to 701)

AUTHORS

Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,

TITLE

Plasmodium yoelii EST project at TIGR

JOURNAL

Unpublished (2001)

COMMENT

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

FEATURES

Source

Email: carltonetigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

Location/Qualifiers
1. 701

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone_1lb="PYB36"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with pY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridAP vector and plasmid DNA isolated."

BASE COUNT 320 a 62 c 147 g 172 t

ORIGIN

Query Match

13.8%; Score 51.6; DB 13; Length 701;

Best Local Similarity 55.6%; Pred. No. 0.00044;

Matches 99; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```

OY 198 TGAAGACAGCTCAACAGAGATGATCAATCGTTGATGCGAGATTGACCA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 512 TGAAGATGATGAAGAGATGAGATGATGTCGAGAGAGATGAAGATGAAGA 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 258 AGAAAAATTTGTTGTCAGCAAACTTCAAGAAATTTTGTACAGACTCGAGGTTAA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 572 GCATGAAGATGACGAGAGAGATGATGATGACGAAGAATGAAGATGAAGAGATGA 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 318 AGATGACAGGTAATCGCGAGTACAGTACAGTGAAGTGAAGTGAAGTGAAGTAA 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 632 AGATGAAGATGATGAAGAGATGATGATGAAGAGATGAAGAGATGAAGATGAA 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

BM165474

LOCUS

DEFINITION BM165474 721 bp mRNA linear EST 04-DEC-2001

EST567997 PYBS Plasmodium yoelii yoelii cDNA clone pYCM68 5' end,

mRNA sequence.

ACCESSION

BM165474

VERSION

BM165474.1 GI:17311155

KEYWORDS

EST.

SOURCE

Plasmodium yoelii yoelii.

ORGANISM

Plasmodium yoelii yoelii.

REFERENCE

1 (bases 1 to 721)

AUTHORS

Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,

TITLE

Plasmodium yoelii EST project at TIGR

JOURNAL

Unpublished (2001)

COMMENT

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Unpublished (2001)
Contact: Jane Carlton

For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

Location/Qualifiers

/organism="Plasmodium yoelli yoelli"
 /strain="17Xt"
 /db_xref="taxon:73239"
 /clone_1lib="PYBS"
 /clone_1lib="PYCLF94"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-28 parasitemia, blood was collected from BALB/cBYJ mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridzAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybridzAP vector and plasmid DNA isolated."

| | | | | |
|------------|-------|------|------|------|
| BASE COUNT | 137 a | 11 c | 75 g | 49 t |
| ORIGIN | | | | |

| | | | | |
|-----------------------|-------|-------------|---------|-------------|
| Query Match | 13.3% | Score 50; | DB 13; | Length 272; |
| Best Local Similarity | 55.1% | Pred. No. | 0.0011; | |
| Matches | 08 | Connections | 0 | |

[illegible]

| LOCUS | 274 bp | mRNA | linear | EST 04-DEC-2001 |
|------------|-----------|------------------------|-------------------|-----------------|
| DEFINITION | EST569409 | pyBS Plasmodium yoelli | yoelli cDNA clone | pyCN214 5' end, |

[illegible]

JOURNAL
COMMENT

JOURNAL COMMENT
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ARCC
<http://www.malaria.mri4.org/mr4pages/index.html>
Seq primer: ADF.
Location/Qualifiers

FEATURES

```

source
1. 276
/organism="Plasmodium yoelli yoelli"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCD34"
/clone_lib="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/vector="pAD-GAL4: At 20-25% parasitemia, blood was
collected from BALB/CBY1 mice infected with PY17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography."
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, eleven
terminal were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZap arms directionally using EcoRI-XhoI cleaved arms
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZap vector and plasmid DNA
isolated."

```

BASE COUNT
ORIGIN

| Query Match | 13.3% | Score 50; | DB 13; | Length 276; |
|-----------------------|--------|---|-----------|-------------|
| Best Local Similarity | 55.1%; | Pred. NO. 0.0011; | | |
| Matches | 98; | Mismatches 80; | Indels 0; | Gaps 0 |
| QY | 198 | TGAAGACCGCTACACAGGAAGTTACTACATGTCGCTGATGGTGAAGATTAGACA | 257 | |
| | | | | |
| Db | 93 | TGAAGATGATGACAGAGATGACATTAAGATGATGACCGAAGAGATGAAGTATGATGAAG | 152 | |
| | | | | |
| QY | 258 | AGAAAAATTTGTTGTCGACGAAACTTCACGGAAAAATTTTTCACAGACTCGGAGGGTAA | 317 | |
| | | | | |
| Db | 153 | GGATGAAGATGACACAGCAAGATCAATATGACCAAGAGATGAAGATGTACGAAAGAAATGA | 212 | |
| | | | | |
| QY | 318 | AGATGACAGGTAAATCGCGCAGGTACAGGTACGAGTCAGTGAAGTTGATGAAGATTTAA | 375 | |
| | | | | |
| Db | 213 | AGATGAAGATGATGAAGAAAGATGAAGTATGATGAAGAGATGAAGAAAGATGAATGA | 270 | |
| | | | | |

| | |
|------------|-----------|
| BM162507 | ACCESSION |
| LOCUS | VERSION |
| DEFINITION | KEYWORDS |
| | SOURCE |
| | ORGANIS |
| | REFERENCE |
| | AUTHORS |
| | TITLE |

| | | | | |
|------------|--|-------------|-------------------------------------|-------------------------------------|
| BM162507 | 425 bp | mRNA | linear | EST 04-DEC-2000 |
| LOCUS | BM162507 | EST565030 | KYBS | Plasmodium yoelli yoelli cDNA clone |
| DEFINITION | EST565030 | KYBS | Plasmodium yoelli yoelli cDNA clone | pyCKR148 5' end, |
| ACCESSION | BM162507 | | | |
| VERSION | BM162507.1 | GI:17308188 | | |
| KEYWORDS | EST. | | | |
| SOURCE | Plasmodium yoelli yoelli. | | | |
| ORGANISM | Plasmodium yoelli yoelli. | | | |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | | |
| AUTHORS | 1 (bases 1 to 425) | | | |
| TITLE | Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Carucci, D.J. | | | |
| | Plasmodium yoelli EST project at TIGR | | | |

Unpublished (2001)
Contact: Jane Carlton

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Location/Qualifiers

1. .425

/organism="Plasmodium yoelli yoelli"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCMT48"
 /clone_lib="PYBS"
 /dev_stage="asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pMD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcytalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridAR arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pMD-GAL4) was excised from the HybridAR vector and plasmid DNA isolated."

| BASE COUNT ORIGIN | 202 a | 27 c | 109 g | 87 t |
|----------------------|-------|------|-------|------|
|----------------------|-------|------|-------|------|

| | | | | |
|-----------------------|--------|-------------------|--------|----------------|
| Query Match | 13.3%; | Score 50; | DB 13; | Length 425; |
| Best Local Similarity | 55.1%; | Pred. No. 0.0011; | | |
| Matches | 98; | Conservative | 0; | Mismatches 80. |
| | | | | Totals |
| | | | | 0. |
| | | | | 0 |

| Conservative | Mismatches | Indels | Gaps |
|--------------|------------|--------|------|
| 0 | 80 | 0 | 0 |

Db
242 TGAAGATGATGAAGAGATGAGATGATGACCGAAGAGATGAAGATGATGAAGA 301

Db 302 GGATGAAGATGACGAAGAGATGACAGTGAAGATGACGAGAGAAAGATGA 361

[illegible]

RESULT 10
BM167376

DEFINITION
EST569899 *PyBS* *Plasmodium yoelli* cDNA clone PYC0128 5' end,
mRNA sequence.

| | | |
|----------|--------------------------|-------------|
| VERSION | BM167376.1 | GI:17300608 |
| KEYWORDS | EST. | |
| SOURCE | Plasmodium yoelii yoelii | |

| ORGANISM | Plasmodium yoelii | Plasmodium yoelii |
|----------|-------------------|-------------------|
| | Plasmodium yoelii | Plasmodium yoelii |

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
1 / haemo 1 +s 453

AUTHORS
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, J.W. Vaidya, S.D.

Fraser, C.M. and Carucci, D.J.
Plasmodium woi 44 500

Unpublished (2001)
Contact: Jane Carlton

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
E-mail: gen1000@igir.org

Location/Qualifiers

1. .453

/organism="Plasmodium yoelli yoelli"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCO128"
 /clone_lib="PYBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4. At 20-28 parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dGTP. After second strand synthesis, uneven termini ligated with pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybrizap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the Hybrizap vector and plasmid DNA isolated."

| | | | | | | | | |
|-------------|-----|---|----|---|-----|---|----|---|
| BASE COUNTY | 223 | a | 25 | c | 112 | g | 93 | t |
| ORIGIN | | | | | | | | |

| | | | | |
|-----------------------|-------|-----------|---------|-------------|
| Query Match | 13.3% | Score 50; | DB 13; | Length 453; |
| Best Local Similarity | 55.1% | Pred. No. | 0.0012; | |
| Matches | 98 | Count | 100 | |

| | | | | | | | | | |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|
| Matches | 98; | Conservative | 0; | Mismatches | 80; | Indels | 0; | Gaps | 0; |
|---------|-----|--------------|----|------------|-----|--------|----|------|----|

198 TGAGACACAGGTCACCAAGGATGTTCTACAAATGTCGTGATGCGAAGATTAGACCA 25
 154 TGAAGTGTGACAGAGATGAAAGATGATGTCAGCATGTGACGACGAAGATGAAAGATGATGAGA 213

Dy
258 AGAAAAATTGTCTGACAGCAAAACTTCACGGAAAATTTTGCACGACTCCGAGGTAA 317

Dc
|| || |
| | | | | | |
214 GGATGAAGATGACAGAAGATGACAGTAGCAGAAGATGTAACATGCCAAGMAGATGA 273

D5
318 AGATGACAGTAAATGCCGCAGCTACAGGTGACCGTCAGATGAAGTTGATGAAGAATTAA 379

D6
274 AGATGAAGATCATGAAGAAGATGAAGATGATGAAGAAGCATCAAGAACAATGAAGATCAA 331

RESULT 11
BM171330

| Accession | Size | Type | Library | EST |
|------------|---|-------------|---------|-----------------|
| U00001 | 459 bp | mRNA | linear | EST 04-DEC-2001 |
| U00002 | 2474/1330 | genomic DNA | clone | pycOT11 5' end, |
| DEFINITION | EST573863 PyBS Plasmodium yoelii cDNA clone pycOT11 5' end, | | | |
| FEATURES | mRNA sequence | | | |

| | |
|-----------|-------------|
| ACCESSION | BM1/1330 |
| VERSION | BM171330.1 |
| KEYWORDS | EST |
| | GI:17304562 |

plasmodium root root
EST.
KEIWORDS
SOURCE

| ORGANISM | Plasmodium yoelii | Plasmodium yoelii |
|----------|-------------------|-------------------|
| | | |

1 (bases 1 to 450)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

AUTHORS
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, J.W., Vaidya, A. & B

Fraser, C.M. and Carucci, D.J.
Plasmodium woiwodi ECM and host cell interaction

JOURNAL COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..459

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCQ11"

/clone_lib="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 225 a 37 c 113 g 84 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 459;

Best Local Similarity 55.1%; Pred. NO. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACAAAGATGTTCTACATGCTGTTGCTGAAGATTAGCA 257

DB 108 TGAAGATGATGAAGAAGATGAAGATGATGACGAAGAAGATGATGAAGA 167

QY 258 AGAAATTTGTTGTCGACGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAA 317

DB 168 GGATGAAGATGACGAAGAAGATGAAGTGAAGATGACGAAGAAGATGA 227

QY 318 AGATGAGGTAAATGCGCAGGTACAGTGAAGTGAAGTGAAGTGAAGTAA 375

DB 228 AGATGAAGATGATGAAGAAGATGAAGATGATGAAGAGATGAAGAAGATGA 285

RESULT 12

BM161175

LOCUS 580 bp mRNA linear EST 04-DEC-2001

DEFINITION EST563698 PYBS Plasmodium yoelii yoelii cDNA clone PYCRB54 5' end,

mRNA sequence.

ACCESSION BM161175.1 GI:17306856

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..580

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCRB54"

/clone_lib="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 251 a 47 c 139 g 143 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 580;

Best Local Similarity 55.1%; Pred. NO. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACAAAGATGTTCTACATGCTGTTGCTGAAGATTAGCA 257

DB 397 TGAAGATGATGAAGAAGATGAAGATGATGACGAAGAAGATGATGAAGA 456

QY 258 AGAAATTTGTTGTCGACGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAA 317

DB 457 GGATGAAGATGACGAAGAAGATGAAGTGAAGATGACGAAGAAGATGA 516

QY 318 AGATGAGGTAAATGCGCAGGTACAGTGAAGTGAAGTGAAGTGAAGTAA 375

DB 517 AGATGAAGATGATGAAGAAGATGAAGATGATGAAGAGATGAAGAAGATGA 574

RESULT 13

BM164042

LOCUS 648 bp mRNA linear EST 04-DEC-2001

DEFINITION EST56565 PYBS Plasmodium yoelii yoelii cDNA clone PYC1P02 5' end,

mRNA sequence.

ACCESSION BM164042.1 GI:17309723

VERSION

KEYWORDS

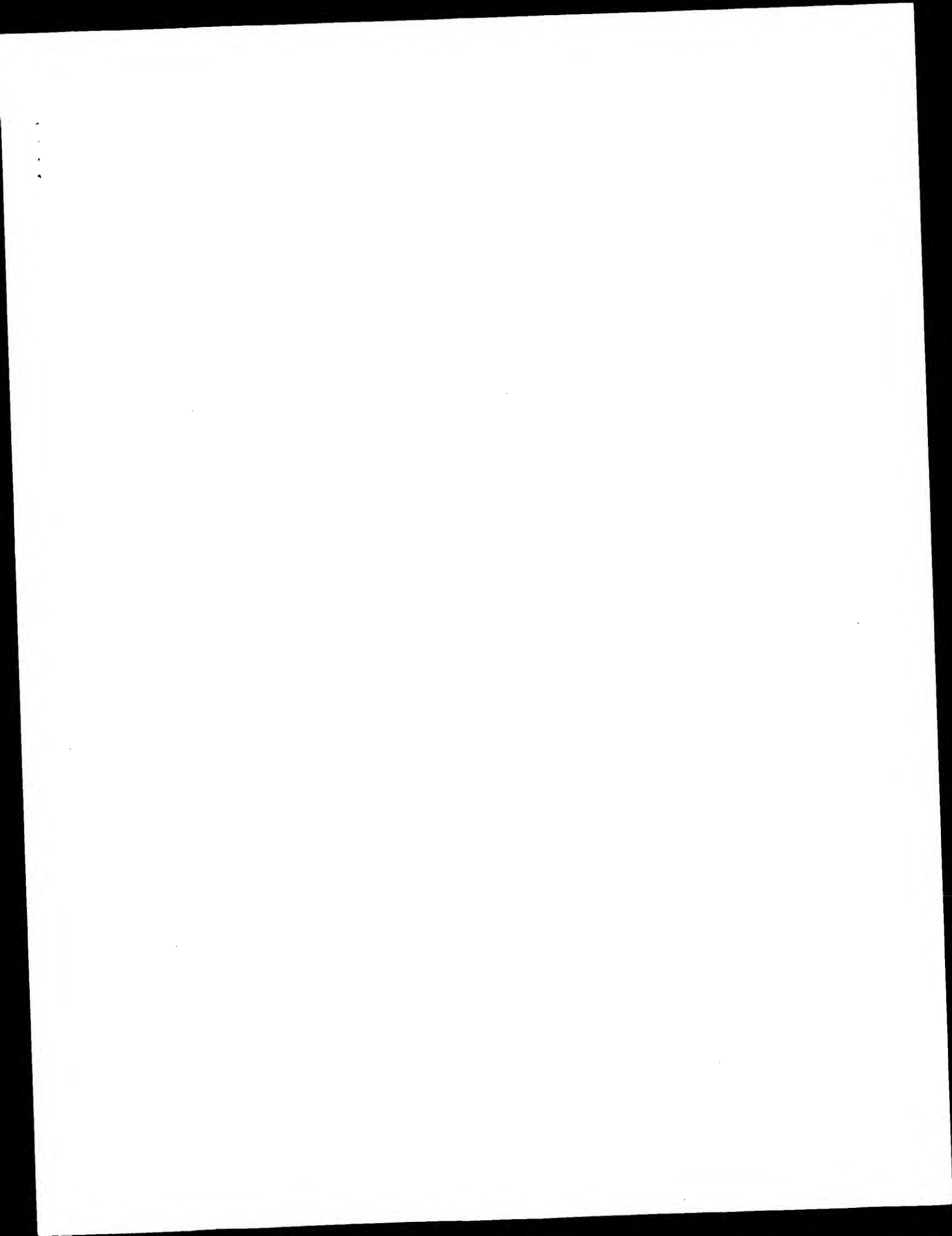
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE



N-PSDB; AAA28180.

57

PT New Saratin polypeptide and gene isolated from Hirudo medicinalis for
 PT blocking platelet adhesion, especially useful for treating or
 PT preventing thrombotic diseases, or for manufacturing a medicament for
 PT thromboembolic diseases

PS Claim 7: Page 41; 46pp; English.

CC Saratin is a protein isolated from the saliva of the medicinal leech
 CC Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
 CC platelet adhesion. The invention includes polynucleotide sequences
 CC encoding Saratin, an expression vector comprising the DNA sequence, a
 CC host cell transformed with the expression vector, antibodies
 CC immunospecific for Saratin, and methods for identifying Saratin agonists
 CC or antagonists. Saratin is useful for treating thromboembolic diseases,
 CC and for manufacturing a medicament for treating thromboembolic diseases.
 CC It is useful for preventing thrombotic diseases. Saratin is also useful
 CC for coating artificial surfaces, since use of Saratin renders them
 CC non-adhesive for cells and prevents the activation of cells. It may also
 CC be used for coating natural collagen surfaces. Furthermore, Saratin is
 CC useful for modifying intracellular lenses in order to lessen the
 CC thrombogenicity of the lens material, for contacting the lens surface, or
 CC for covalent crosslinking to modify the lens material. The lens material
 CC is used for refractive anterior or posterior chamber ocular implants,
 CC which may be implanted into the eye. This new type of coating avoids
 CC problems contributed by stimulated cell growth. In combination with other
 CC medicaments that are for instance conferring cell death, Saratin coating
 CC helps to completely overcome posterior capsule opacification. The
 CC antibody immunospecific for Saratin, as well as Saratin itself, are
 CC useful for measuring samples derived from host cell cultures or from a
 CC treated subject. The present sequence represents the amino acid sequence
 CC of the Saratin protein.

CC SQ Sequence 103 AA:

Query Match 100.0%; Score 573; DB 21; Length 103;
 Best Local Similarity 100.0%; Pred. No. 2.3e-52; Indels 0; Gaps 0;
 Matches 103; Conservative 0; Mismatches 0

OY 1 EEREDCFTYANKRYTDFKSFKSSDLDECKKTCFTEYCYIVEDTVNKECIYNVDS 60
 DB 1 EEREDCFTYANKRYTDFKSFKSSDLDECKKTCFTEYCYIVEDTVNKECIYNVDS 60
 OY 61 EELDQKFFVNDENTENTLTDCGKDGAGNAAGTDESDEVDD 103
 DB 61 EELDQKFFVNDENTENTLTDCGKDGAGNAAGTDESDEVDD 103

RESULT 2

AAR22950 ID AAR22950 standard; Protein: 147 AA.

XX AAR22950;

XX 10-SEP-1992 (first entry)

XX Leech antiplatelet protein.

XX LAPP; collagen; platelet aggregation; antithrombotic; anticoagulant;

KW coronary artery disease; cerebrovascular disease; ss.

XX Haementeria officinalis.

XX Location/Qualifiers

XX Key 1..21

XX Peptide /label= leader peptide amino acids"

XX Protein /note= "21 hydrophobic 22..147

XX /label= mature protein

XX EPA80651-A.

XX 15-APR-1992.

PF 07-OCT-1991; 91EP-0309157.

XX 09-OCT-1990; 90US-0594917.

XX (MERI) MERCK & CO INC.

XX Connolly TN, Keller PM;

XX WPI: 1992-125288/16.

DR N-PSDB; AAQ23682.

PT New protein to inhibit collagen stimulated platelet aggregation -
 PT extracted from salivary gland of Haementeria officinalis for
 PT treating thrombotic diseases

PS Claim 6: Page 15; 20pp; English.

CC When cloned into the yeast expression vector pKH42, the LAPP gene
 CC sequence forms a fusion product of MA α -alpha-1 leader sequence
 CC plus mature LAPP, which when produced are proteolytically processed by
 CC the Lys-Arg cleaving endopeptidase (KEX2), which cleaves on the
 CC C terminal side of Lys-Arg, and the products secreted into the culture
 CC medium.
 CC LAPP blocks stimulation of platelet aggregation by collagen, at an
 CC antithrombotic concentration of about 100 nmolar (or 1.6 ug/ml).
 CC platelet aggregation is stimulated by 2 ug/ml collagen and addition
 CC of LAPP inhibits this with an IC50 of 45nM, but this may be
 CC overcome by the addition of 0.25 mM arachidonic acid.
 CC LAPP could be used for treatment and prevention of thrombotic
 CC conditions especially coronary artery and cerebrovascular disease
 CC in mammals.
 CC See also AAQ23871-7.

CC SQ Sequence 147 AA:

Query Match 14.0%; Score 80.5; DB 13; Length 147;
 Best Local Similarity 32.7%; Pred. No. 1.1; Indels 3; Gaps 1;
 Matches 18; Conservative 10; Mismatches 24

OY 5 DCWTYANKRYTDFKSFKSSDLDECKKTCFTEY---CYIVEDTVNKECIYN 56
 DB 66 DCWKRPGWKLPDNLITFTSTVDECKKCEESAVPSCYIIQINTETNECYRN 120

RESULT 3

AAW68010 ID AAW68010 standard; Protein: 412 AA.

XX AAW68010;

XX 27-APR-1999 (first entry)

XX S. frugiperda immunophilin FKBP46.

XX Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;

KW transplant; tissue graft.

XX Spodoptera frugiperda.

XX Key 111..114

XX Domain /note= "EEAP motif"

XX Region /note= "putative nuclear localisation signal"

XX Region /note= "putative nuclear localisation signal"

XX Domain /note= "putative nuclear localisation signal"

XX Domain /note= "AP motif"

XX Domain /note= "EEAP motif"

XX region /note= "putative nuclear localisation signal"

us-09-936-737a-2.rag

```

XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-061415O.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI, 2001-656860/75.
XX DR N-PsDB; ABL13970.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 36393; 21pp + Sequence Listing; English.
XX XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL01675) and the encoded proteins
CC (ABW57737-ABW72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pcl_sequences.
XX XX Sequence 368 AA;
SQ
Query Match 13.4%; Score 77; DB 22; Length 368;
Best Local Similarity 23.9%; Pred. No. 7.8; Indels 34; Gaps 6;
Matches 28; Conservative
QY 14 KYTDFPKS-----FKSSDDECKTCFCFYCYIFE--DTVKNKCTYNVDG 60
    :|||:|:| | | | | | | | | | | | | | | | | | | | |
DB 89 EYTDADRSVAPTEFLAEFLSAMVKDYKA---LKYCKLLIQYEPNATKEFPYL-- 143
    ||:|:| | | | | | | | | | | | | | | | | | | | |
QY 61 ELDDQEFNV----DENFTEN-----YLTDCGKDGNGAGTGDESDEVDED 103
    |:|:| | | | | | | | | | | | | | | | | | | | |
DB 144 --LDKLRAVATSSDPDENYKSSSPDLALDLHNSVDYADVDSDEGADDEGDADAD 148
    ||:|:| | | | | | | | | | | | | | | | | | | | |

RESULT 5
AAB14260
ID AAB14260 standard; Protein: 418 AA.
XX AC AAB14260;
XX DT 09-FEB-2001 (first entry)
XX DE Mouse Ice-4 protein sequence #2.
XX CD ccd-3; virally induced cell death; apoptosis; gene therapy; neural;
KM muscular degenerative disease; myocardial infarction; stroke; aging;
KW interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
XX OS Mus sp.
XX PN US6083735-A.
XX PD 04-JUL-2000.
XX PF 10-JUN-1994; 94US-0258287.
XX PR 24-JUN-1993; 93US-0080850.
XX PA (GEHO ) GEN HOSPITAL CORP.
XX PI Yuan J, Miura M;
XX
```

DR WPI: 2000-464343/40.

XX New human Ich-1L and Ich-1S proteins for negative and positive
PT regulation of programmed cell death and for developing therapeutic
PT methods for diseases and conditions characterized by cell death, e.g.
XX myocardial infarction or stroke
PS Disclosure; Fig 17; 121pp; English.

XX The present sequence is a mouse Ice-4 protein sequence. The present
CC sequence was used in a sequence homology comparison with the present
CC Interleukin-1beta converting enzyme (ICE) (AAB14253), murine
CC (AAB14252) and C. elegans ced-3 (AAB14246). The coding sequence of the
CC cell death (apoptosis). Ich-1 may play an important role in both the
CC positive and negative regulation of apoptosis. The Ich gene may be used
CC in gene therapy in disorders characterized by cell death e.g. neural and
CC muscular degenerative diseases, myocardial infarction, stroke, vitally
XX induced cell death and aging.

SQ Sequence 418 AA;

Query Match

Best Local Similarity 13.2%; Score 75.5; DB 21; Length 418;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DQKRVVD--ENFTENYL--TDCEGR-DAGNAGTG 94

DB 26 VFDDIVER---NVLNGDELTKIGESASPTLNKAKNLVENFLERTDMAGKIFAGHIANSQ 81

OY 95 DE-----SDEVD 101

DB 82 EQLSLQPSNDEDD 94

RESULT 6

AA550109

ID AAY50109 standard; Protein; 419 AA.

AC AAY50109;

DT 21-JAN-2000 (first entry)

DE Murine caspase-12 splice variant, caspase-12L.

XX Caspase; splice variant; truncated; programmed cell death; apoptosis;
KW regulation; proteolytic cascade; malignant condition; cancer;
KW chromic lymphocytic leukemia; prostatic hypertrophy;
KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
ds.

OS Mus sp.

XX Key Location/Qualifiers

FT Cleavage-site 94..95

FT Protein 95..419

FT Cleavage-site 318..319

PN W09952925-A1.

PD 21-OCT-1999.

PF 14-APR-1999; 99WO-US08064.

PR 16-APR-1998; 98US-0081962.

PA (GEHO) GEN HOSPITAL CORP.

PI Yuan J, Morishima N.

DR WPI: 1999-620369/53.

DR N-PSDB: AAZ32645.

XX New nucleic acid encoding the short form of caspase-12, used e.g. for
PT treating tumors
XX

PS Claim 1; Fig 1; 68pp; English.

XX This sequence represents a cDNA encoding murine caspase-12L, a splice
CC variant of caspase-12. Caspases are a family of proteins involved in
CC the regulation of apoptosis and are synthesized as proforms which are
CC activated via cleavage after specific Asp residues. Mammalian cells
CC express several caspases, and it is thought that these act in a
CC proteolytic cascade to cause programmed cell death. Nucleic acids
CC encoding caspase-12S (AA32644) or truncated forms of caspase-12L
CC (AA32646, AA32647) are used for production, recombinantly or in vivo,
CC particularly useful for treating (pre)malignant conditions (e.g. solid
CC tumours, B cell lymphoma, chronic lymphocytic leukemia, prostatic
CC hyper trophy, preneoplastic liver foci and resistance to chemotherapy), or
CC autoimmune diseases. The caspase-12 proteins can also be used to raise
CC specific antibodies (for example, to determine gene expression and to
CC screen expression libraries) or as molecular weight markers. Fragments
CC of caspase-12 encoding nucleic acids can be used as probes to isolate
CC hybridisation for chromosomal location of the caspase-12 gene, and for
CC Northern blotting to determine caspase-12 mRNA expression in tissues.

SQ Sequence 419 AA;

Query Match

Best Local Similarity 13.2%; Score 75.5; DB 20; Length 419;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DQKRVVD--ENFTENYL--TDCEGR-DAGNAGTG 94

DB 26 VFDDIVER---NVLNGDELTKIGESASPTLNKAKNLVENFLERTDMAGKIFAGHIANSQ 81

OY 95 DE-----SDEVD 101

DB 82 EQLSLQPSNDEDD 94

RESULT 7

AAB14259

ID AAB14259 standard; Protein; 419 AA.

AC AAB14259;

DT 09-FEB-2001 (first entry)

DE Mouse Ice-4 protein sequence #1.

XX ced-3; virally induced cell death; apoptosis; gene therapy; neural;
KW muscular degenerative disease; myocardial infarction; stroke; aging;
KW Interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
OS Mus sp.

PN US6083735-A.

PD 04-JUL-2000.

PF 10-JUN-1994; 94US-0258287.

PR 24-JUN-1993; 93US-0080850.

PA (GEHO) GEN HOSPITAL CORP.

PI Yuan J, Miura M.

DR WPI: 2000-464343/40.

DR N-PSDB: AAZ72843.

CC The present invention is a caspase-mediated apoptosis. These include the
CC useful for modulating caspase-12 protein shown here. These agents can be used
CC caspase-12 protein shown here.

| Matches | 26; Conservative | 27; Liberal |
|---------|------------------|-------------|
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 7 | 7 |
| 8 | 8 | 8 |
| 9 | 9 | 9 |
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12 |
| 13 | 13 | 13 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |
| 21 | 21 | 21 |
| 22 | 22 | 22 |
| 23 | 23 | 23 |
| 24 | 24 | 24 |
| 25 | 25 | 25 |
| 26 | 26 | 26 |
| 27 | 27 | 27 |
| 28 | 28 | 28 |
| 29 | 29 | 29 |
| 30 | 30 | 30 |
| 31 | 31 | 31 |
| 32 | 32 | 32 |
| 33 | 33 | 33 |
| 34 | 34 | 34 |
| 35 | 35 | 35 |
| 36 | 36 | 36 |
| 37 | 37 | 37 |
| 38 | 38 | 38 |
| 39 | 39 | 39 |
| 40 | 40 | 40 |
| 41 | 41 | 41 |
| 42 | 42 | 42 |
| 43 | 43 | 43 |
| 44 | 44 | 44 |
| 45 | 45 | 45 |
| 46 | 46 | 46 |
| 47 | 47 | 47 |
| 48 | 48 | 48 |
| 49 | 49 | 49 |
| 50 | 50 | 50 |
| 51 | 51 | 51 |
| 52 | 52 | 52 |
| 53 | 53 | 53 |
| 54 | 54 | 54 |
| 55 | 55 | 55 |
| 56 | 56 | 56 |
| 57 | 57 | 57 |
| 58 | 58 | 58 |
| 59 | 59 | 59 |
| 60 | 60 | 60 |
| 61 | 61 | 61 |
| 62 | 62 | 62 |
| 63 | 63 | 63 |
| 64 | 64 | 64 |
| 65 | 65 | 65 |
| 66 | 66 | 66 |
| 67 | 67 | 67 |
| 68 | 68 | 68 |
| 69 | 69 | 69 |
| 70 | 70 | 70 |
| 71 | 71 | 71 |
| 72 | 72 | 72 |
| 73 | 73 | 73 |
| 74 | 74 | 74 |
| 75 | 75 | 75 |
| 76 | 76 | 76 |
| 77 | 77 | 77 |
| 78 | 78 | 78 |
| 79 | 79 | 79 |
| 80 | 80 | 80 |
| 81 | 81 | 81 |
| 82 | 82 | 82 |
| 83 | 83 | 83 |
| 84 | 84 | 84 |
| 85 | 85 | 85 |
| 86 | 86 | 86 |
| 87 | 87 | 87 |
| 88 | 88 | 88 |
| 89 | 89 | 89 |
| 90 | 90 | 90 |
| 91 | 91 | 91 |
| 92 | 92 | 92 |
| 93 | 93 | 93 |
| 94 | 94 | 94 |
| 95 | 95 | 95 |
| 96 | 96 | 96 |
| 97 | 97 | 97 |
| 98 | 98 | 98 |
| 99 | 99 | 99 |
| 100 | 100 | 100 |

QY 44 VFEDTVNKECYNVVDGEEL-----DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 40 VFDDLYER-----NVNGBELKIGESASFILNKAEINVENFLEKTDMAKTIYFAGHIANSO 95
 QY 95 DE-----SDEVD 101
 DB 96 EQLSLQFSNDEDD 108

RESULT 10
 AAR98464
 ID AAR98464 standard; Protein; 432 AA.
 AC AAR98464;
 XX
 DT 25-SEP-1996 (first entry)
 DE Murine Ich-3.
 DE
 XX
 KW Ich-3; murine ICE-ced-3 homologue; programmed cell death;
 KM apoptosis; interleukin-1 beta converting enzyme; gene therapy.
 XX
 OS Mus sp.
 PN WO9620721-A1.
 PD 11-JUL-1996.
 PF 04-JAN-1996; 96MO-US00177.
 PR 04-JAN-1995; 95US-0368704.
 PA (GEMO) GEN HOSPITAL CORP.
 PI Miura M, Yuan J;
 PS WPI: 1996-333763/33.
 DR N-PSDB: AAT31554.
 XX

Preventing or promoting programmed cell death in vertebrate cells -
 PT comprises inhibiting or increasing the activity of
 PT interleukin-1 beta converting enzyme, or altering expression of
 PT other related genes
 PS Claim 24: Fig 14; 127pp; English.
 XX
 CC Ich-3 (AAR98464) causes programmed cell death and shows significant
 CC homology to mouse interleukin-1 beta converting enzyme (ICE),
 CC mouse mitch-2 (AAR98461) and human Ich-1 (AAR98462-63). Its sequence
 CC was deduced from the Ich-3 gene (AAT31554) isolated from a mouse thymus
 CC cDNA library. The protein can be obtained from host cells containing
 CC vectors that include an Ich-3 coding sequence. It can be used to
 CC control the programmed cell death of vertebrate cells, to develop cell
 CC lines that remain viable for extended periods, and to increase the
 CC activity of ICE.
 XX
 SQ Sequence 432 AA;

Query Match 13.2%; Score 75.5; DB 17; Length 432;
 Best Local Similarity 35.6%; Pred. No. 14;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 QY 44 VFEDTVNKECYNVVDGEEL-----DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 40 VFDDLYER-----NVNGBELKIGESASFILNKAEINVENFLEKTDMAKTIYFAGHIANSO 95
 QY 95 DE-----SDEVD 101
 DB 96 EQLSLQFSNDEDD 108

RESULT 11
 ABB66916

ID ABB66916 standard; Protein; 1154 AA.
 XX
 AC ABB66916;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 27540.
 DE
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001MO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL11019.
 XX

New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure: SEQ ID NO 27540; 21pp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB101840-AB16175) and the encoded proteins
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1154 AA;

Query Match 12.8%; Score 73.5; DB 22; Length 1154;
 Best Local Similarity 28.1%; Pred. No. 73;
 Matches 27; Conservative 14; Mismatches 42; Indels 13; Gaps 4;
 QY 12 NRYTDFDKSFKSSDLDCEK-KTCTFTEYCYIVFEDTVNKECYN-VYDGEELDQEFV 69
 DB 505 SKRDIKDFKSSMDLDDKLFKFAELVRCMFLMKRMOQDLKGNKVCNVEKDNVWQ 564
 QY 70 VDNEFTENYL--TDCEGDA-----GNAAGTG 94
 DB 565 IVENFWRHYICCCQCKKDKACALSNPSMLGNTERTYTG 600

RESULT 12
 AAG17789
 ID AAG17789 standard; Protein; 175 AA.
 AC AAG17789;
 XX
 DT 17-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 18943.
 DE
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130649.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
04-MAY-1999; 99US-0132407.
05-MAY-1999; 99US-0132484.
06-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135553.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139761.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148319.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149175.
17-AUG-1999; 99US-0149368.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151066.
30-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0153363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157173.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.5%; Score 71.5; DB 21; Length 175;
Best Local Similarity 28.6%; Pred. No. 12;

Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;

QY 23 KSSDLDECKKTKCTEYIVFDVTKKCYNNVVDGEEL-DOEKRVNDENFTENLT- 80
DB 93 KKS-----KRALPEESFTWTFDQHK-----DAGDEIHVADIIEDLMSNPITY 141

QY 81 -----DCEGDAGNAAGTGDSEDEVDED 103
DB 142 FNNDADEEDFDGDDGDGEDEDDDEED 172

RESULT 13
AAG17788
ID AAG17788 standard; Protein; 256 AA.

XX AAG17788;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 18942.

XX Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.
XX EPI03405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139495.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.


```

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144337.
PR 21-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```

```

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 12.5%; Score 71.5; DB 21; Length 256;
Best Local Similarity 28.6%; Pred. No. 19;
Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;

```

QY 23 KKSDDLECKKTCFTECYIYFEDTVNKECYNNVDEEL-DOEKFYVDENFTENTYLT-80
DB 174 KKS-----KRALPESFFFTFDQKKE-----DAGDEIHVDADIKEDLWSNPLYT 222
QY 81 -----DOEGKDGAGNAAGTGESDEVED 103
DB 223 FNNDADEDFDGDGDGDEGEDDDDEED 253

```

RESULT 14
AA18263
ID AAB18263 standard; Protein: 320 AA.

AA18263;
XX
DT 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:120.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.

OS Plasmodium falciparum.
XX
PN W0200025728-A2.
XX

PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99MO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI: 2000-365347/31.
 DR
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 PS Disclosure: Page 283-284; 577pp; English.
 XX
 PS The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasite lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 320 AA;
 Query Match 12.5%; Score 71.5; DB 21; Length 320;
 Best Local Similarity 27.8%; Pred. No. 25;
 Matches 27; Conservative 9; Mismatches 24; Indels 37; Gaps 5;
 YY 9 EVANRRY-----TDEPKS-----FKKSDLECKR-----TCFKTEYC----- 41
 DB 51 FYALKRYEKSLEKKRDEKSNNDKISIKSYKDFKNELQIITIKNEYCIGCEGIITNY 110
 YY 42 ---YVEEDTVNKECYNVVDEELDOEKFEVVDENET 75
 DB 111 DEYIITIEYKENDSL-----KFEDEVFVLDKNVT 140
 RESULT 15
 ID AAB95086 standard; Protein; 589 AA.
 XX
 AC AAB95086;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:16999.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.

XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 16999; 2537pp + CD ROM; English.
 XX
 PS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 589 AA;
 Query Match 12.4%; Score 71; DB 22; Length 589;
 Best Local Similarity 23.8%; Pred. No. 59;
 Matches 30; Conservative 23; Mismatches 37; Indels 36; Gaps 6;
 YY 12 NRKTTDDPKSEKFS-----SDLECKRTT-----CFKTE---ICYIVE 46
 DB 158 SKEFTQNNKREKKNIVYHTDSSLEKORTLDGTSSEIKYPRICSTREKMSVQLI 217
 YY 47 DTVNKECYNVVDEELDOEKFEVVD-----EFTFNLYLTDECGKADAGNAGTGDSE 97
 DB 218 MTRSDGYENSTDEGMDKDALEDESESVSEIGSESENETTSY-GRASGDGDSSEDE 276
 YY 98 DEVDPD 103
 DB 277 EE-DDD 281

Search completed: February 25, 2003, 02:32:13
 Job time : 28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:31:10 ; Search time 15 Seconds

(without alignments)
202.037 Million cell updates/sec

Title: US-09-936-737A-2

Perfect score: 573

Sequence: 1 EFERDCWTFYANRKYTPDK.....GKDAQNAAGTDSDEVD 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | ID | Description |
|------------|-------|---------|--------------|-----------------------|--------------------|
| 1 | 80.5 | 14.0 | 126 | 1 US-08-044-547-3 | Sequence 3, Appli |
| 2 | 80.5 | 14.0 | 147 | 1 US-08-044-547-1 | Sequence 1, Appli |
| 3 | 80.5 | 14.0 | 412 | 2 US-08-741-134-2 | Sequence 2, Appli |
| 4 | 75.5 | 13.2 | 172 | 4 US-09-187-789-16 | Sequence 16, Appli |
| 5 | 75.5 | 13.2 | 172 | 4 US-09-139-600-11 | Sequence 11, Appli |
| 6 | 75.5 | 13.2 | 418 | 3 US-08-258-287B-58 | Sequence 58, Appli |
| 7 | 75.5 | 13.2 | 418 | 3 US-08-358-704C-56 | Sequence 56, Appli |
| 8 | 75.5 | 13.2 | 419 | 3 US-08-258-287B-57 | Sequence 57, Appli |
| 9 | 75.5 | 13.2 | 419 | 3 US-08-358-704C-55 | Sequence 55, Appli |
| 10 | 69.5 | 12.1 | 124 | 1 US-08-170-360-1 | Sequence 1, Appli |
| 11 | 69.5 | 12.1 | 674 | 4 US-08-961-083-200 | Sequence 200, App |
| 12 | 67 | 11.7 | 124 | 1 US-08-170-360-2 | Sequence 2, Appli |
| 13 | 67 | 11.7 | 124 | 1 US-08-888-497-41 | Sequence 41, Appli |
| 14 | 67 | 11.7 | 124 | 4 US-09-352-230-41 | Sequence 41, Appli |
| 15 | 67 | 11.7 | 124 | 5 PCT-US94-07926-41 | Sequence 41, Appli |
| 16 | 67 | 11.7 | 146 | 4 US-08-888-497-34 | Sequence 34, Appli |
| 17 | 67 | 11.7 | 146 | 4 US-09-352-230-34 | Sequence 34, Appli |
| 18 | 67 | 11.7 | 146 | 4 PCT-US94-07926-34 | Sequence 34, Appli |
| 19 | 67 | 11.7 | 299 | 4 US-09-059-023-20 | Sequence 20, Appli |
| 20 | 66.5 | 11.6 | 86 | 4 US-09-134-001C-3986 | Sequence 3986, Ap |
| 21 | 66.5 | 11.6 | 2182 | 2 US-08-487-826B-16 | Sequence 16, Appli |
| 22 | 65.5 | 11.3 | 725 | 1 US-08-813-940-25 | Sequence 25, Appli |
| 23 | 64.5 | 11.3 | 126 | 1 US-08-170-360-3 | Sequence 3, Appli |
| 24 | 64.5 | 11.3 | 126 | 4 US-08-888-497-38 | Sequence 38, Appli |
| 25 | 64.5 | 11.3 | 126 | 4 US-09-352-230-38 | Sequence 38, Appli |
| 26 | 64.5 | 11.3 | 126 | 5 PCT-US94-07926-38 | Sequence 38, Appli |
| 27 | 64.5 | 11.3 | 148 | 2 US-08-888-497-36 | Sequence 36, Appli |

| | | | | | |
|----|------|------|------|---------------------|--------------------|
| 28 | 64.5 | 11.3 | 148 | 4 US-09-352-230-36 | Sequence 36, Appli |
| 29 | 64.5 | 11.3 | 148 | 5 PCT-US94-07926-36 | Sequence 36, Appli |
| 30 | 64 | 11.2 | 519 | 4 US-09-172-841-55 | Sequence 55, Appli |
| 31 | 63 | 11.0 | 294 | 4 US-09-355-160D-8 | Sequence 8, Appli |
| 32 | 63 | 11.0 | 1658 | 4 US-08-609-049A-13 | Sequence 13, Appli |
| 33 | 63 | 11.0 | 1658 | 4 US-09-170-996-13 | Sequence 13, Appli |
| 34 | 63 | 11.0 | 1726 | 2 US-08-609-049A-30 | Sequence 30, Appli |
| 35 | 63 | 11.0 | 1726 | 4 US-09-170-996-30 | Sequence 30, Appli |
| 36 | 62.5 | 10.9 | 250 | 2 US-08-861-269-5 | Sequence 5, Appli |
| 37 | 62.5 | 10.9 | 250 | 2 US-09-134-596-5 | Sequence 5, Appli |
| 38 | 62.5 | 10.9 | 250 | 3 US-09-293-273-5 | Sequence 19, Appli |
| 39 | 62.5 | 10.9 | 1172 | 1 US-08-313-288B-19 | Sequence 19, Appli |
| 40 | 62.5 | 10.9 | 2555 | 3 US-09-058-489-36 | Sequence 36, Appli |
| 41 | 61.5 | 10.7 | 346 | 2 US-08-702-153-2 | Sequence 2, Appli |
| 42 | 61.5 | 10.7 | 346 | 2 US-08-702-153-4 | Sequence 4, Appli |
| 43 | 61.5 | 10.7 | 416 | 2 US-09-211-930-3 | Sequence 3, Appli |
| 44 | 61.5 | 10.7 | 416 | 3 US-09-340-993-3 | Sequence 3, Appli |
| 45 | 61.5 | 10.7 | 416 | 4 US-09-468-442-3 | Sequence 3, Appli |

ALIGNMENTS

```

RESULT 1
US-08-044-547-3
Sequence 3, Application US/08044547
Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
TITLE OF INVENTION: Protein for Inhibiting
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Merck & Co., Inc.
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-044-547-3
Query Match 14.0%; Score 80.5; DB 1; Length 126;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

```


Wed Feb 26 15:59:00 2003

us-09-936-737a-2.ra1

Page 3

Db 26 VFDDLVK-----NVINGDELLKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81
QY 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 5
US-09-139-600-11
Sequence 11, Application US/09139600
Patent No. 6432628

GENERAL INFORMATION:
APPLICANT: Alnemert, Emad S.
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
FILE REFERENCE: 480140.434
CURRENT FILING DATE: 1998-08-25
CURRENT APPLICATION NUMBER: US/09/139,600
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 11
LENGTH: 172
TYPE: PRT
ORGANISM: Mus musculus
US-09-139-600-11

Query Match 13.2%; Score 75.5; DB 4; Length 172;
Best Local Similarity 35.6%; Pred. No. 1;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVKKECYNVVDGEEL---DDEKRVVD--ENFTENTYL--TDCEGK-DAGNAGTG 94
26 VFDDLVK-----NVINGDELLKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81
Db 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 6
US-08-258-287B-58
Sequence 58, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 24636 SSK
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-258-287B-58

Query Match 13.2%; Score 75.5; DB 3; Length 418;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVKKECYNVVDGEEL---DDEKRVVD--ENFTENTYL--TDCEGK-DAGNAGTG 94
26 VFDDLVK-----NVINGDELLKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81
Db 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 7
US-08-368-704C-56
Sequence 56, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-368-704C-56

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DOEKFYVD--ENFTENYL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELTKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 81
OY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 8

US-08-258-287B-57
Sequence 57, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609,3920001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-258-287B-57

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DOEKFYVD--ENFTENYL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELTKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 81
OY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 9
US-08-368-704C-55

Sequence 55, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:

APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 4-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609,3920002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-368-704C-55

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DOEKFYVD--ENFTENYL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELTKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 81
OY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 10

US-08-170-360-1
Sequence 1, Application US/08170360
Patent No. 5656602
GENERAL INFORMATION:
APPLICANT: Tseong, Albert P. S.
TITLE OF INVENTION: PL2 INHIBITORY COMPOUNDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg Ernst & Kurz
STREET: Suite 701-E, 555 Thirteenth St., N.W
CITY: Washington
STATE: D. C.

COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-1

Query Match 12.1%; Score 69.5; DB 1; Length 124;
Best Local Similarity 27.1%; Pred. No. 2.9;
Matches 16; Conservative 8; Mismatches 20; Indels 15; Gaps 1;

QY 28 LDECKTKCTEYCYIVEDTVNKECYVNVVGEELDQEFVNDENFTENTDCEKRD 86
Db 38 VDELRCETHDNCY-----RDAKNLDSCKFLVNPFTESISCSNTE 81

RESULT 11
US-08-961-083-200
Sequence 200, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match 12.1%; Score 69.5; DB 4; Length 674;
Best Local Similarity 23.6%; Pred. No. 23;
Matches 25; Conservative 17; Mismatches 35; Indels 29; Gaps 5;

QY 7 WTFYANKRYTDEKSFKSSDLDECKT-----CFTEYCYIVFEDTVNKECYVNVVGE 61
Db 322 WIKOLNRLKRPD-----SEILNCRYTSEQISCYKSDISYVF-----LRYFIADFV 368

QY 62 ELQGEF-----VVDENFTENTDCEG-----KDGNAAGTGDE 96
Db 369 QEDKALYLDGDIWVTKNLDLFAFDLQDYPPLAVRDFGGRAYFGOE 414

RESULT 12
US-08-170-360-2
Sequence 2, Application US/08170360
Patent No. 5656602
GENERAL INFORMATION:
APPLICANT: Tseng, Albert P. S.
TITLE OF INVENTION: PLAZ INHIBITORY COMPOUNDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg Ernst & Kurz
STREET: Suite 701-E, 555 Thirteenth St., N.W
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-2

Query Match 11.7%; Score 67; DB 1; Length 124;

Best Local Similarity 22.7%; Pred. No. 5.4; Indels 22; Gaps 3;
Matches 17; Conservative 14; Mismatches 22

Db 40 DLDRC---CQTHHCY-----NQAKLESCKFLIDNPYNTYSYKC----- 77

Qy 27 DLDECKKTCFKTEYCIYFEDTVNKECYNVVDEGEILDQKFFVVDENFTENYLTDCGKD 86
40 DLDRC---CQTHHCY-----NQAKLESCKFLIDNPYNTYSYKC----- 77

Qy 87 AGNAGTGSDEVD 101
Db 78 SGNVITCSDKNDCE 92

RESULT 13
US-08-888-497-41
Sequence 41, Application US/08888497
Patent No. 5972677
GENERAL INFORMATION:
APPLICANT: Tischfield, Jay A.
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
STREET: 200 East Broward Boulevard
CITY: Fort Lauderdale
STATE: FL
COUNTRY: USA
ZIP: 33301

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/888,497
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,405
FILING DATE:
APPLICATION NUMBER: US 08/097,354
FILING DATE: 26-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: IN21044-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305-527-2498
TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-888-497-41

Query Match 11.7%; Score 67; DB 2; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

Qy 27 DLDECKKTCFKTEYCIYFEDTVNKECYNVVDEGEILDQKFFVVDENFTENYLTDCGKD 86

Db 40 DLDRC---CQTHHCY-----NQAKLESCKFLIDNPYNTYSYKC----- 77
Qy 87 AGNAGTGSDEVD 101
Db 78 SGNVITCSDKNDCE 92

RESULT 14
US-09-362-230-41
Sequence 41, Application US/09362230
Patent No. 6352849

GENERAL INFORMATION:
APPLICANT: Tischfield, Jay A.
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
STREET: 200 East Broward Boulevard
CITY: Fort Lauderdale
STATE: FL
COUNTRY: USA
ZIP: 33301

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/362,230
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/888,497
FILING DATE:
APPLICATION NUMBER: US 08/097,354
FILING DATE: 26-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: IN21044-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305-527-2498
TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-362-230-41

Query Match 11.7%; Score 67; DB 4; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

Qy 27 DLDECKKTCFKTEYCIYFEDTVNKECYNVVDEGEILDQKFFVVDENFTENYLTDCGKD 86
40 DLDRC---CQTHHCY-----NQAKLESCKFLIDNPYNTYSYKC----- 77
Qy 87 AGNAGTGSDEVD 101
Db 78 SGNVITCSDKNDCE 92

RESULT 15
PCT-US94-07926-41


```

: Sequence 41, Application PC/TUS9407926
: GENERAL INFORMATION:
: APPLICANT: Tischfield, Jay A.
: APPLICANT: Seilhamer, Jeffrey J.
: TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
: TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
: TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
: STREET: 200 East Broward Boulevard
: CITY: Fort Lauderdale
: STATE: FL
: COUNTRY: USA
: ZIP: 33301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/07926
: FILING DATE: 15-JUL-1994
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/097,354
: FILING DATE: 26-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Manso, Peter J.
: REGISTRATION NUMBER: 32,264
: REFERENCE/DOCKET NUMBER: IN21044-5
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 305-527-2498
: TELEFAX: 305-764-4996
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 124 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-07926-41

Query Match      11.7%; Score 67; DB 5; Length 124;
Best Local Similarity 22.7%; Pred. NO. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKKTCFTEYCYIVFEDTVNKECYVNVVDEELDQKRFVVDENFTENYLTDCBGKD 86
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 DLDRG---CQTHDRY-----NOAKKLESCKFLIDNPFTINTYSKRC----- 77
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 87 ACNAGTDESDVD 101
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 SGNVITCSDKNNDC 92

```

Search completed: February 25, 2003, 02:33:47
 Job time : 16 secs


```

; APPLICANT: Glimcher, Laurie R.
; TITLE OF INVENTION: Phospholipase A2 Group Preferentially
; FILE REFERENCE: HUT-046
; CURRENT APPLICATION NUMBER: US/09/993,999
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 60/246,316
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-993-999-7

```

```

Query Match
Best Local Similarity 11.9%; Score 68; DB 10; Length 146;
Matches 15; Conservative 9; Mismatches 18; Indels 18; Gaps 2;

```

```

OY 27 DLDECKKTCFTEKYCYIVFEDTVNKECYNVVDGEELDOEFVVDENFTENYLTDCGKRD 86
DB 62 DLDRCC---CQTHRCY-----SOAKLSECKFLDNPFTNTYISYSCSGSE 103

```

```

RESULT 6
US-09-801-574-57
; Sequence 57, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 2789
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-801-574-57

```

```

Query Match
Best Local Similarity 11.6%; Score 66.5; DB 10; Length 2789;
Matches 25; Conservative 13; Mismatches 23; Indels 35; Gaps 6;

```

```

OY 11 ANRKYTDPRK-----SPKSSDLDCKKTC-----FKTECYIVFEDTVNKECYNVVDG 60
DB 1349 SSKSYLDRKRLTYDSFASSTYPHCEOSCKREKELKTEGC-----SSGNCILH--TDG 1399
OY 61 EELDOEFVVDENFTENYLTDCGKDGAGNAAGTDE 96
DB 1400 NE-----TWTENYELDV-----ASTEDD 1419

```

```

RESULT 7
US-09-974-298-114
; Sequence 114, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10

```

```

; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Inocyte ID No. US20020156263A1 3267030C01
US-09-974-298-114

```

```

Query Match
Best Local Similarity 11.3%; Score 65; DB 9; Length 1125;
Matches 24; Conservative 24; Mismatches 39; Indels 18; Gaps 4;

```

```

OY 1 EEREDCWFYANRKYTDPRKSFKSSDLDCKKTCFTEKYCYIVFEDTVNKECYNVVDG 60
DB 1032 EEDD-----SEKEEEDKEMELQEEKCEKPCQDEE-----EEEEEE-----VEE 1075
OY 61 EELDOEFVVDENFTENYLTDCGKDGAGNAAG--TGDESDVEDD 103
DB 1076 EEVEEAENEGEAEKTEGIMKDRASQASSLQKYGESSEQVSSB 1120

```

```

RESULT 8
US-09-925-297-511
; Sequence 511, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 511
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (156)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-511

```

```

Query Match
Best Local Similarity 11.3%; Score 64.5; DB 10; Length 156;
Matches 17; Conservative 5; Mismatches 20; Indels 15; Gaps 2;

```

```

OY 28 LDECKKTCFTEKYCYIVFEDTVNKECYNVVDGEELDOEFVVDENFTENYLTDCG 84
DB 68 VDBLDRCCQTHRCY-----DQAK-----LDSCFKFLDNPFTNTYISYSCSG 109

```

```

RESULT 9
US-10-042-417-52
; Sequence 52, Application US/10042417
; Patent No. US20020123082A1
; GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; FILE REFERENCE: 5914-090-999
; CURRENT APPLICATION NUMBER: US/10/042,417
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,179
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 89

```

Wed Feb 26 15:59:01 2003

us-09-936-737a-2.rapb

Page 4

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 52
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: All Xaa positions
; OTHER INFORMATION: xaa-unknown amino acid residue
US-10-042-417-52

Query Match
Best Local Similarity 27.1%; Pred. No. 45;
Matches 23; Conservative 14; Mismatches 38; Indels 10; Gaps 4;

OY 16 TDFPKSKSSDDECKKTCFTECYIVFEDTVNKECYIVD---GEEIDQ---KP 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 TSHSGELKTS-TSKITSTAMKMK--DITMSTKQYACLDLTKNGIGEEIDNHPWTKP 312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 69 VVDENFTENYITDCEGDAGNAGT 93
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 VSENFETSPYWMLEADLADIEDT 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-10-092-219-8
; Sequence 8, Application US/10092219
; Patent No. US2002015114A1
; GENERAL INFORMATION:
; APPLICANT: Domlin, Jan
; TITLE OF INVENTION: No. US2002015114A1el Lipid Kinase
; FILE REFERENCE: 1064HG/50947
; CURRENT APPLICATION NUMBER: US/10/092,219
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: PCY/GB98/00244
; PRIOR FILING DATE: 1998-01-27
; PRIOR APPLICATION NUMBER: 09/355,160
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 9701652.1
; PRIOR FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-219-8

Query Match
Best Local Similarity 11.0%; Score 63; DB 12; Length 294;
Matches 19; Conservative 18; Mismatches 25; Indels 4; Gaps 2;

OY 12 NRRYDPEKSKSSDDECKKTCFTECYIVFEDTVNKECYIVD---GEEIDQ---BKF 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45 DROWDPSRGEFELSSSTQROGFNKKQDYDLWFPPELDSOKRAVD-IDVEKLTQAELEKI 103
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 69 VVDENF 74
    | | | | |
Db 104 LLDDNF 109
    | | | | |

RESULT 11
US-09-930-871-8
; Sequence 8, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; APPLICANT: Mathur, Brian
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleot
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
```

```

; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1381
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1381)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-930-871-8

Query Match
Best Local Similarity 11.0%; Score 63; DB 10; Length 1381;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

OY 10 YANRKYTDF-DKSF-KKSSDDECKKTCFTECYIVFEDTVNKE--CYIVVD--GEEI 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1014 YKRRIYEFIQOSFTRKQKILDEIKP-----LDDLNNKSCSMNHTXKIGKDL 1062
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 64 DOEKFVVDENFTENYITDCEGDAGNAGTGDSEDEVD 102
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1063 D-----YKDVNGTTSIGIGSSVEKYIIDE 1068
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-930-871-10
; Sequence 10, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; APPLICANT: Mathur, Brian
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1387
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1387)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-930-871-10

Query Match
Best Local Similarity 11.0%; Score 63; DB 10; Length 1387;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

OY 10 YANRKYTDF-DKSF-KKSSDDECKKTCFTECYIVFEDTVNKE--CYIVVD--GEEI 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1014 YKRRIYEFIQOSFTRKQKILDEIKP-----LDDLNNKSCSMNHTXKIGKDL 1062
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 64 DOEKFVVDENFTENYITDCEGDAGNAGTGDSEDEVD 102
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1063 D-----YKDVNGTTSIGIGSSVEKYIIDE 1068
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-930-871-18
; Sequence 16, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 1392
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1392)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-930-871-18

```

```

Query Match          11.0%; Score 63; DB 10; Length 1392;
Best Local Similarity 28.3%; Pred. No. 1.5e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

```

```

QY 10 YANKRYTDF-DKSF-KKSSDLDECKKTCFTECYIVFEDTVNKE--CYNVVD--GEEL 63
Db 1025 YVKKRIYEFIQGSFIRKQKILDEIKP-----LDDLNNKKSCKSMNHTXIEIGKDL 1073
QY 64 DOEKFVVDENFTENYLTDCGKDGAGNAAGTGDESDVEDE 102
Db 1074 D-----YLKDVNGTTSIGIGTSSVEKIIDE 1099

```

```

RESULT 14
US-09-930-871-20
; Sequence 20, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1398)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-930-871-20

```

```

Query Match          11.0%; Score 63; DB 10; Length 1398;
Best Local Similarity 28.3%; Pred. No. 1.5e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

```

```

QY 10 YANKRYTDF-DKSF-KKSSDLDECKKTCFTECYIVFEDTVNKE--CYNVVD--GEEL 63
Db 1025 YVKKRIYEFIQGSFIRKQKILDEIKP-----LDDLNNKKSCKSMNHTXIEIGKDL 1073
QY 64 DOEKFVVDENFTENYLTDCGKDGAGNAAGTGDESDVEDE 102
Db 1074 D-----YLKDVNGTTSIGIGTSSVEKIIDE 1099

```

```

RESULT 15
US-09-930-871-6
; Sequence 6, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; APPLICANT: Mathur, Brian
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1442
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1442)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-930-871-6

```

```

Query Match          11.0%; Score 63; DB 10; Length 1442;
Best Local Similarity 28.3%; Pred. No. 1.6e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

```

```

QY 10 YANKRYTDF-DKSF-KKSSDLDECKKTCFTECYIVFEDTVNKE--CYNVVD--GEEL 63
Db 1014 YVKKRIYEFIQGSFIRKQKILDEIKP-----LDDLNNKKSCKSMNHTXIEIGKDL 1062
QY 64 DOEKFVVDENFTENYLTDCGKDGAGNAAGTGDESDVEDE 102
Db 1063 D-----YLKDVNGTTSIGIGTSSVEKIIDE 1088

```

Search completed: February 25, 2003, 02:34:14
Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:32:55 ; Search time 2615 Seconds
(without alignments)
1146.306 Million cell updates/sec

Title: US-09-936-737A-2
Perfect score: 573
Sequence: 1 EEREDCWTRYANKRYTDFDK.....GMDAGNAGTGEDEDEVED 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_plus_p2n.model -DEV-xlh
-Q/cg92.1/USPTO_SPOOL/US09936737/runat.14022003.100732.2291/app.query.fasta.1.263
-DB=GenEmbl -QFWT=fastcap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936737/cgcn.1.1.2425/etunat.14022003.100732.2291 -NCPD=6 -ICPU=3
-NO_XIPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : GenEmbl:
1: gb_da:
2: gb_htg:
3: gb_in:
4: gb_ov:
5: gb_ov:
6: gb_ov:
7: gb_ov:
8: gb_ov:
9: gb_ov:
10: gb_ov:
11: gb_ov:
12: gb_ov:
13: gb_ov:
14: gb_ov:
15: gb_ov:
16: gb_ov:
17: gb_ov:
18: gb_ov:
19: gb_ov:
20: gb_ov:
21: gb_ov:
22: gb_ov:
23: gb_ov:
24: gb_ov:
25: gb_ov:
26: gb_ov:
27: gb_ov:
28: gb_ov:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------|---------------------|
| 1 | 573 | 100.0 | 375 | 6 AX036541 | AX036541 Sequence |
| 2 | 90 | 15.7 | 110000 | 2 PFMAL13P2_1 | Continuation (2 of |
| 3 | 90 | 15.7 | 110000 | 2 PFMAL13P2_2 | Continuation (3 of |
| 4 | 88 | 15.4 | 138024 | 2 AC131381 | AC131381 Strongylo |
| 5 | 86.5 | 15.1 | 128960 | 9 HSJ919F19 | AL109947 Human DNA |
| 6 | 85.5 | 14.9 | 563 | 9 HEI1APP | M81489 Haemeteria |
| 7 | 83 | 14.5 | 95615 | 9 AL138835 | AL138835 Human DNA |
| 8 | 83 | 14.5 | 149109 | 2 AC101018 | AC101018 Rattus no |
| 9 | 83 | 14.5 | 159492 | 2 AC115173 | AC115173 Rattus no |
| 10 | 81.5 | 14.2 | 2255 | 6 AR030964 | U15038 Spodoptera |
| 11 | 81.5 | 14.2 | 2255 | 6 AR030964 | AK030964 Sequence |
| 12 | 80.5 | 14.0 | 178681 | 2 PFDMACPT | X84041 P.falciparu |
| 13 | 80.5 | 13.9 | 110000 | 2 AC098520 | AC098520 Rattus no |
| 14 | 79.5 | 13.9 | 232370 | 2 AC108272_2 | Continuation (3 of |
| 15 | 79.5 | 13.9 | 232370 | 2 AC108272_2 | Continuation (3 of |
| 16 | 79 | 13.8 | 43380 | 3 CER26D4A | AL110478 Caenorhab |
| 17 | 78.5 | 13.7 | 7311 | 1 SPN131985 | AT131985 Streptococ |
| 18 | 78.5 | 13.7 | 32110 | 1 CER40G12 | Z77661 Caenorhabdi |
| 19 | 78.5 | 13.7 | 212085 | 10 AL663032 | AL663032 Mouse DNA |
| 20 | 78.5 | 13.7 | 301550 | 1 AP003192 | AP003192 Clostridi |
| 21 | 78 | 13.6 | 44237 | 3 AC024847 | X64346 Herpesvirus |
| 22 | 78 | 13.6 | 112830 | 14 HSGEND | AL034558 Herpesv |
| 23 | 78 | 13.6 | 158548 | 3 PFMAL13P2 | AL035616 Homo sapi |
| 24 | 78 | 13.6 | 174280 | 2 AL356116 | AC006889 Caenorhab |
| 25 | 78 | 13.6 | 267118 | 2 AC006889 | AC006889 Caenorhab |
| 26 | 78 | 13.6 | 294250 | 1 AP001517 | AP001517 Bacillus |
| 27 | 78 | 13.6 | 318221 | 2 PFMAL13P3 | AL049184 Plasmodiu |
| 28 | 77.5 | 13.5 | 43570 | 8 AB019231 | AB019231 Arabidops |
| 29 | 77.5 | 13.5 | 46341 | 9 AL445568 | AL445568 Human DNA |
| 30 | 77.5 | 13.5 | 199531 | 2 AC006281 | AC006281 Plasmodiu |
| 31 | 77 | 13.4 | 500 | 3 AC006281 | M55686 T.brucei ac |
| 32 | 77 | 13.4 | 2625 | 3 TRRCARML | AY071572 Drosophil |
| 33 | 77 | 13.4 | 141040 | 8 AP004382 | AP004382 Oryza sat |
| 34 | 76.5 | 13.4 | 34960 | 8 U58762 | U58762 Caenorhabdi |
| 35 | 76 | 13.3 | 2252 | 3 U58762 | U20549 Rastrepia m |
| 36 | 76 | 13.3 | 35766 | 3 CBMG47C05 | CBMG47C05 Caenorhab |
| 37 | 76 | 13.3 | 37640 | 3 CER27C14 | Z68135 Caenorhabdi |
| 38 | 76 | 13.3 | 121290 | 2 AC111055 | Z68135 Caenorhabdi |
| 39 | 76 | 13.3 | 159001 | 2 AC012139 | AC012139 Homo sapi |
| 40 | 76 | 13.3 | 164748 | 9 AC068672 | AC068672 Homo sapi |
| 41 | 76 | 13.3 | 19583 | 9 AC007305 | AC007305 Mus muscu |
| 42 | 76 | 13.3 | 214839 | 6 AR102119 | AR102119 Sequence |
| 43 | 75.5 | 13.2 | 1402 | 6 AR103163 | AR103163 Sequence |
| 44 | 75.5 | 13.2 | 1402 | 6 AR103163 | BC028979 Mus muscu |
| 45 | 75.5 | 13.2 | 2228 | 10 BC028979 | BC028979 Mus muscu |

RESULT 1

ALIGNMENTS

| BASE COUNT | ORIGIN |
|------------|--------|
| 123 a | 62 c |
| | 85 g |
| | 105 t |

| | |
|------------------------|----------|
| Alignment Scores: | |
| Prod. No.: | 4.21e-60 |
| Score: | 573.00 |
| Percent Similarity: | 100.00% |
| Best Local Similarity: | 100.00% |
| Query Match: | 100.00% |
| DB: | 6 |
| Length: | 375 |
| Matches: | 103 |
| Conservative: | 0 |
| Mismatches: | 0 |
| Indels: | 0 |
| Gaps: | 0 |

US-09-936-737A-2 (1-103) x AX036541 (1-375)

| | | | |
|----|-----|---|-----|
| QY | 1 | GIUGIARIGLuspcysrtrphrhrPhetylalaaanaqlvstYrThrAspPheasplys | 20 |
| Db | 64 | GAGAACGTGAAACATTGTGGACGTTTTACGGGAACAATAATACAGACTTGATAAA | 122 |
| QY | 21 | SeepHeLysLysSerSerAspLeuaspGluCysLysLysThrCysPheLysThhGluTyr | 40 |
| Db | 124 | TCTTTTAAAGAACTCCTGATCTTGATGACCAATGCCAAAAAATGTTCAAGACGGAAGTAC | 183 |
| QY | 41 | CysTyrIleValPheGluAspPheAlaInslysgLcysTyrTyrAsnaValAspGly | 60 |
| Db | 184 | TGCTACATCGTTTTGGAAGACACGGCTCAACAAGGAATGTAACTCAATGCTGTGAATGT | 243 |
| QY | 61 | GluGluLeuAspGluGluLysPheValValAspGluAsnPhetrrGluAsnTyrLeuThr | 80 |
| Db | 244 | GAAAGGTTAAGCAAGAAAAATTGTGTGCGACGAAACCTTCAACGGAAATTTATTGACA | 303 |
| QY | 81 | AspCysGluGlyLysAspAlaGluAspAlaAlaGlyThrGluLyspGluSerAspGluVal | 100 |
| Db | 304 | GACTCGGAGGTAAAGATGCCAGGTAATCGCGACGGTACAGGTGACGATCGATGAAGTTT | 363 |
| QY | 101 | AspGluasp 103 | |
| Db | 364 | GATGAGAGAT 372 | |

| | | | | |
|---------------------------------|---------------|--------|-----------|----------|
| RESULT 2 | PFMAL13P2_1/c | | | |
| WPCOMMENT | | | | |
| Sequence split into 4 fragments | | | | |
| Fragment Name | Begin | End | Accession | AL049185 |
| PFMAL13P2_0 | 1 | 110000 | | |

Alignment Scores:

| Prod No. | Score |
|----------|-------|
| 1 | 100 |
| 2 | 100 |
| 3 | 100 |
| 4 | 100 |
| 5 | 100 |
| 6 | 100 |
| 7 | 100 |
| 8 | 100 |
| 9 | 100 |
| 10 | 100 |
| 11 | 100 |
| 12 | 100 |
| 13 | 100 |
| 14 | 100 |
| 15 | 100 |
| 16 | 100 |
| 17 | 100 |
| 18 | 100 |
| 19 | 100 |
| 20 | 100 |
| 21 | 100 |
| 22 | 100 |
| 23 | 100 |
| 24 | 100 |
| 25 | 100 |
| 26 | 100 |
| 27 | 100 |
| 28 | 100 |
| 29 | 100 |
| 30 | 100 |
| 31 | 100 |
| 32 | 100 |
| 33 | 100 |
| 34 | 100 |
| 35 | 100 |
| 36 | 100 |
| 37 | 100 |
| 38 | 100 |
| 39 | 100 |
| 40 | 100 |
| 41 | 100 |
| 42 | 100 |
| 43 | 100 |
| 44 | 100 |
| 45 | 100 |
| 46 | 100 |
| 47 | 100 |
| 48 | 100 |
| 49 | 100 |
| 50 | 100 |
| 51 | 100 |
| 52 | 100 |
| 53 | 100 |
| 54 | 100 |
| 55 | 100 |
| 56 | 100 |
| 57 | 100 |
| 58 | 100 |
| 59 | 100 |
| 60 | 100 |
| 61 | 100 |
| 62 | 100 |
| 63 | 100 |
| 64 | 100 |
| 65 | 100 |
| 66 | 100 |
| 67 | 100 |
| 68 | 100 |
| 69 | 100 |
| 70 | 100 |
| 71 | 100 |
| 72 | 100 |
| 73 | 100 |
| 74 | 100 |
| 75 | 100 |
| 76 | 100 |
| 77 | 100 |
| 78 | 100 |
| 79 | 100 |
| 80 | 100 |
| 81 | 100 |
| 82 | 100 |
| 83 | 100 |
| 84 | 100 |
| 85 | 100 |
| 86 | 100 |
| 87 | 100 |
| 88 | 100 |
| 89 | 100 |
| 90 | 100 |
| 91 | 100 |
| 92 | 100 |
| 93 | 100 |
| 94 | 100 |
| 95 | 100 |
| 96 | 100 |
| 97 | 100 |
| 98 | 100 |
| 99 | 100 |
| 100 | 100 |

US-09-936-737A-2 (1-103) x PFMALL3P2_1 (1-110000)

0v 18 -----Phosphatase 101438
GAGGGTCTCCATCAAAAAAGCAGAAATTACGTGATTTATTAATAAAAAAATAATAA

QY 31 Cys-----LysLysThrCysPheLysThrGluTyrCysTwrTleValPheCln-----Phe 47

48 ThrValAsnLysGluCysTyrTrpAsnValVal----- 58

Db 101257 GTTCTAAATCAAACCTTCAAAAAATAATTATTTGCTTTGGAGT----- 58

Db 101197 ACTACGCAAGATTATCTGATGTGAAGAATCGTCACCAATAGATGATTTCATACTATT 101138

-----GATATACTTCAGAT 101093

RESULT 3
EMAL13P2 2/C

| Fragment Name | Begin | End |
|---------------|-------|--------|
| PMAL13P2_0 | 1 | 110000 |

Alignment Scores:

| | | |
|---------------|--------|--------|
| Length: | 33.0 | 110000 |
| Matches: | 90.00 | 33 |
| Conservative: | 34.64% | 20 |
| Mismatches: | 21.57% | 40 |
| Indels: | 15.71% | 60 |
| Gaps: | | 2 |

5-09-936-737A-2 (1-103) x FEMAL13P2_2 (1-110000)

5 AspCysTirPThrPheIYrAlaAsnArlGlySTyrThrAsp-----17

10 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100 102 104 106 108 110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 188 190 192 194 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232 234 236 238 240 242 244 246 248 250 252 254 256 258 260 262 264 266 268 270 272 274 276 278 280 282 284 286 288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318 320 322 324 326 328 330 332 334 336 338 340 342 344 346 348 350 352 354 356 358 360 362 364 366 368 370 372 374 376 378 380 382 384 386 388 390 392 394 396 398 400 402 404 406 408 410 412 414 416 418 420 422 424 426 428 430 432 434 436 438 440 442 444 446 448 450 452 454 456 458 460 462 464 466 468 470 472 474 476 478 480 482 484 486 488 490 492 494 496 498 500 502 504 506 508 510 512 514 516 518 520 522 524 526 528 530 532 534 536 538 540 542 544 546 548 550 552 554 556 558 560 562 564 566 568 570 572 574 576 578 580 582 584 586 588 590 592 594 596 598 600 602 604 606 608 610 612 614 616 618 620 622 624 626 628 630 632 634 636 638 640 642 644 646 648 650 652 654 656 658 660 662 664 666 668 670 672 674 676 678 680 682 684 686 688 690 692 694 696 698 700 702 704 706 708 710 712 714 716 718 720 722 724 726 728 730 732 734 736 738 740 742 744 746 748 750 752 754 756 758 760 762 764 766 768 770 772 774 776 778 780 782 784 786 788 790 792 794 796 798 800 802 804 806 808 810 812 814 816 818 820 822 824 826 828 830 832 834 836 838 840 842 844 846 848 850 852 854 856 858 860 862 864 866 868 870 872 874 876 878 880 882 884 886 888 890 892 894 896 898 900 902 904 906 908 910 912 914 916 918 920 922 924 926 928 930 932 934 936 938 940 942 944 946 948 950 952 954 956 958 960 962 964 966 968 970 972 974 976 978 980 982 984 986 988 990 992 994 996 998 1000

18 -----pheaspLysSerPheLysSerSerasp-----LeuaspGlu 30

```

Db 1437 TTACATTTGATATAAATTCACAAATAATTCAGATCATGATTAACCTATATATGAA 1378
Oy 31 Cys-----LysLysThrCysPheLysThrGluTyrCysTyrIleValPheGlu---Asp 47
Db 1377 TGCATATATAAATAATGATATACACAAAAGAAATATTTTCCCTTGATATGATTTAA 1318
Oy 48 ThrValAsnLysGluCysTyrTyrAsnValVal----- 58
Db 1317 GTATATAGTAAAGTCTGATTTATATATTTATCAATGAGATATAAATTAAGAGCT 1258
Oy 58 ----- 58
Db 1257 GTTCTAATCAACCTCAAAAAAATAATTTAGATTTGATATCATGATGAGACAA 1198
Oy 59 -----AspGlyGluGluLeuAspGluGluValPheVal 69
Db 1197 ACTAAGCAGCAATATCTGATGTAAGAAAGTCTCACAATATGATGATTTCAATCTATT 1138
Oy 70 ValAspGluAsnPheThrGluAsnTyrLeuThrAspCysGluGluLysAspAlaGlyAsn 89
Db 1137 GTTGATATAAATAATATATCAGATGATTA-----GATATATCTTCAGAT 1093
Oy 90 AlaAlaGlyThrGlyAspGluSerAspGluValAspGlu 102
Db 1092 GAAGCTACAGAAAAAGACGAAAAAGATCAAGTGTATGAA 1054
RESULT 4
AC131381/c 138024 bp DNA linear HTG 21-AUG-2002
LOCUS Strongylocentrotus purpuratus clone Sp31J8, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 ordered pieces.
AC131381 GI:22380705
VERSION HTG: HTGS_PHASE2: HTGS_FULMOP.
KEYWORDS Strongylocentrotus purpuratus.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinodermata; Echinodermata; Echinodermata;
Strongylocentrotus; Strongylocentrotus.
REFERENCE
AUTHORS Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C.,
Yun, C.-H., Minkawa, T., Amore, G., Homan, V., Arenas-Mena, C.,
Olm, O., Brown, C., Titus, L., C.B., Lee, P.Y., Revilla, R.,
Schlstra, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z., Arnone, M.I.,
Rowen, L., Cameron, R., Andrew, M., McClay, D.R., Hood, L. and Bolouri, H.
A provisional regulatory gene network for specification of
endomesoderm in the sea urchin embryo
Dev. Biol. 246 (1), 162-190 (2002)
TITLE
JOURNAL
MEDLINE 22024134
REFERENCE 2 (bases 1 to 138024)
AUTHORS Rowen, L., Cameron, R.A. and Davidson, E.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2002) Multimegabase Sequencing Center, Institute
for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA
COMMENT
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: ISBMS
Web site: http://www.systemsbio.org
Contact: leerowen@systemsbiology.org
Drafting center: ISBMS
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990395

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 58994: contig of 58994 bp in length
* 58995 59094: gap of unknown length
* 59095 96228: contig of 37134 bp in length
* 96229 96329: gap of unknown length
* 96329 138024: contig of 41696 bp in length.
FEATURES
Source
Location/Qualifiers
1..138024
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Sp31J8"
/clone_lib="Caltech Strongylocentrotus purpuratus sperm
genomic BAC library A"
/note="This library is described in Cameron, R.A.,
Maharaj, G., Rast, J.P., Martinez, P., Blom, T.R.,
Swartzell, S., Wallace, J.C., Pouzet, A.J., Livingston,
B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J.,
Davidson, E.H., Hood, L. Proc. Natl. Acad. Sci. U.S.A. 97
(17): 9514-8 2000"
BASE COUNT 42761 a 25096 c 25960 g 43507 t 700 others
ORIGIN
Alignment Scores:
Pred. No.: 76.2 Length: 138024
Score: 88.0 Matches: 18
Percent Similarity: 52.83% Conservative: 10
Best Local Similarity: 33.96% Mismatches: 7
Query Match: 15.36% Indels: 18
DB: 2 Gaps: 2
US-09-936-737a-2 (1-103) x AC131381 (1-138024)
Oy 31 CysLysLysThr-----CysPheLysThrGlu 39
Db 133186 TGCAGCCGACACGACATCTAGACACAGCATGCTGTCAGTGGCTTAAACATCA 133127
Oy 40 TyrCysTyrIleValPheGluAspPheValAsnLysGluCysTyrTyrAsnValVal--- 58
Db 133126 TACTGTTTCATATATTTTGAACACCTCTTAATTAAGATTAATTAATCTCTCATAT 133067
Oy 59 -----AspGlyGluGluLeuAspGlu 65
Db 133066 GACACATCTGACGACCAATGGAAGATCTTGATAG 133028
RESULT 5
HSJ919F19/c 128960 bp DNA linear PRI 12-JUN-2001
LOCUS Human DNA sequence from clone RP5-919F19 on chromosome 6q16.3-22.1,
DEFINITION complete sequence.
ACCESSION AL109947
VERSION AL109947.19 GI:13897493
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Johnson, C.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 30, 2001 this sequence version replaced gi:13751877.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em', EMBL, SW', SWISSPROT, Tr', TREMBL, Wp', WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP5-919P19 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm> VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-919P19 The true left end of clone RP5-1036D8 is at 86977 in this sequence. The true right end of clone RP11-425D10 is at 56374 in this sequence.

FEATURES

Source

Location/Qualifiers
1..128960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q16.3-22.1"
/clone="RP5-919P19"
/clone_lib="RPCI-5"
16..330
repeat_region
/note="AluY repeat: matches 1..298 of consensus"
878..1177
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
1409..1704
repeat_region
/note="AluX repeat: matches 1..296 of consensus"
2306..2424
repeat_region
/note="FLAM_A repeat: matches 1..123 of consensus"
3935..4225
repeat_region
/note="AluSg repeat: matches 1..299 of consensus"
4512..4698
repeat_region
/note="L1MC5 repeat: matches 7347..7537 of consensus"
5185..5506
repeat_region
/note="AluDo repeat: matches 1..308 of consensus"
5558..6086
repeat_region
/note="L1MC5 repeat: matches 6834..7364 of consensus"
6137..6330
repeat_region
/note="L1MC/D repeat: matches 5276..5467 of consensus"
6393..6686
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
6998..7112
repeat_region
/note="FLAM_A repeat: matches 1..115 of consensus"
7186..7588
repeat_region
/note="HAI1 repeat: matches 505..914 of consensus"
7996..8355
repeat_region
/note="L1MA2 repeat: matches 3613..3976 of consensus"
8372..8434
repeat_region
/note="L1MA2 repeat: matches 4133..4193 of consensus"
8435..8737
repeat_region
/note="AluX repeat: matches 1..298 of consensus"
8738..9144
repeat_region
/note="L1MA2 repeat: matches 4193..4614 of consensus"
9339..11100
repeat_region
/note="MER66-internal repeat: matches 4895..6676 of consensus"
11101..11496
repeat_region
/note="MER66B repeat: matches 1..392 of consensus"
11510..12032
repeat_region
/note="L1MA2 repeat: matches 4604..5114 of consensus"
12033..12338
repeat_region
/note="AluSg repeat: matches 1..306 of consensus"
12339..13410
repeat_region
/note="L1MA2 repeat: matches 5114..6307 of consensus"
13633..13852
repeat_region

repeat_region
/note="AluDb repeat: matches 87..307 of consensus"
13896..15355
repeat_region
/note="RIGER2 repeat: matches 1266..2718 of consensus"
15356..15657
repeat_region
/note="AluX repeat: matches 1..301 of consensus"
15658..16040
repeat_region
/note="RIGER2 repeat: matches 863..1266 of consensus"
16041..16350
repeat_region
/note="AluDo repeat: matches 1..306 of consensus"
16351..16417
repeat_region
/note="RIGER2 repeat: matches 798..863 of consensus"
16445..16538
repeat_region
/note="FLAM_C repeat: matches 24..117 of consensus"
16579..16625
repeat_region
/note="RIGER2 repeat: matches 1..47 of consensus"
16749..17174
repeat_region
/note="MER65A repeat: matches 1..445 of consensus"
17696..17996
repeat_region
/note="AluX repeat: matches 1..306 of consensus"
18497..18587
repeat_region
/note="FLAM_A repeat: matches 43..130 of consensus"
18596..18897
repeat_region
/note="AluSg repeat: matches 1..302 of consensus"
19209..19517
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
20539..20659
repeat_region
/note="MIR repeat: matches 98..228 of consensus"
20912..21038
repeat_region
/note="AluDo repeat: matches 1..126 of consensus"
21039..21330
repeat_region
/note="AluSg repeat: matches 1..291 of consensus"
21331..21467
repeat_region
/note="AluDo repeat: matches 126..266 of consensus"
21473..21756
repeat_region
/note="AluDo repeat: matches 14..295 of consensus"
22240..22524
repeat_region
/note="AluX repeat: matches 1..286 of consensus"
23046..23585
repeat_region
/note="L12 repeat: matches 2234..2745 of consensus"
23679..24214
repeat_region
/note="MER41A repeat: matches 1..538 of consensus"
24274..24440
repeat_region
/note="MIR repeat: matches 17..191 of consensus"
24622..24927
repeat_region
/note="AluX repeat: matches 1..305 of consensus"
26536..26843
repeat_region
/note="AluDb repeat: matches 1..303 of consensus"
27436..27726
repeat_region
/note="AluDb repeat: matches 13..304 of consensus"
29235..29339
repeat_region
/note="L12 repeat: matches 2596..2705 of consensus"
29588..29879
repeat_region
/note="AluSc repeat: matches 1..293 of consensus"
30927..31173
repeat_region
/note="L1R23 repeat: matches 204..437 of consensus"
31638..31785
repeat_region
/note="MER5A repeat: matches 87..189 of consensus"
31786..31914
repeat_region
/note="FLAM_C repeat: matches 1..131 of consensus"
31915..31969
repeat_region
/note="MER5A repeat: matches 2..87 of consensus"
32211..32511
repeat_region
/note="AluX repeat: matches 1..301 of consensus"
33285..33395
repeat_region
/note="MER81 repeat: matches 3..114 of consensus"
33429..33566
repeat_region
/note="FLAM_C repeat: matches 1..138 of consensus"
33567..33606
repeat_region
/note="20 copies 2 mer tg 87% conserved"
34452..34750
repeat_region
/note="AluSc repeat: matches 1..299 of consensus"
34798..35089
repeat_region
/note="AluDb repeat: matches 1..292 of consensus"

```

repeat_region      35591..35905
                    /note="AluX repeat: matches 1..312 of consensus"
repeat_region      35917..35978
                    /note="L2 repeat: matches 2622..2650 of consensus"
repeat_region      36099..37024
                    /note="AluX repeat: matches 1..311 of consensus"
repeat_region      37215..37349
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
repeat_region      37451..37586
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
repeat_region      37718..37742
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
repeat_region      37743..38041
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
repeat_region      38042..38272
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
repeat_region      38504..38786
                    /note="L1M4S repeat: matches 6146..6295 of consensus"
misc_feature       39348..40556
                    /note="CPG island"
misc_feature       /evidence=not_experimental
                    complement(42531..43073)
                    /note="match: STS: Em:636500"
repeat_region      43639..45763
                    /note="L2 repeat: matches 2617..2750 of consensus"
repeat_region      47289..47314
                    /note="L3 copies 2 mer ca 100% conserved"
repeat_region      47413..47506
                    /note="L1M2 repeat: matches 128..244 of consensus"

```

Alignment Scores:

| Pred. No.: | 107 | Length: | 128960 |
|------------------------|--------|---------------|--------|
| Score: | 86.50 | Matches: | 32 |
| Percent Similarity: | 42.11% | Conservative: | 16 |
| Best Local Similarity: | 28.07% | Mismatches: | 49 |
| Query Match: | 15.10% | Indels: | 17 |
| DB: | 9 | Gaps: | 5 |

US-09-936-737a-2 (1-103) x HSJ919F19 (1-128960)

```

QY 3 ArgGluAspCysTrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPhe 22
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115992 CACGCTGACAAAAAGTGTCTTAAACACCCCAAGTGTGAGATATTAAACCTTAT 115933

QY 23 LysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys-----ThrGluTyr 40
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115932 TGGGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115873

QY 41 -----CysTyrL1LevalPheGluAspThrValAsnLys 51
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115872 TCCAGATGATTTGTTATAGTGTCTTTGTTACATATTAT-----ATGTTATATAA 115819

QY 52 GluCysTyrTyr-----AsnValVal---AspGluGluGluLeuAspGluGluLys 67
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115818 AGGGAATATTTTCTTCTTGAATGTTTGAAGATGATGATGATGATGATGATGATGATGAT 115759

QY 68 PheValValAspGluAspLeuThrGluAsnTyrLeuThrAspCysGluGluLysAspAla 87
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115758 CTGAGAGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115699

QY 88 GlyAsnAlaAlaGlyThrGlyAspGluSerAspGluValAsp 101
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115698 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115657

```

RESULT 6

LOCUS HEILAPP 563 bp mRNA linear INV 26-APR-1993
 DEFINITION Haementeria officialis anti-platelet protein (LAPP) mRNA, complete cds.
 ACCESSION M81489
 VERSION M81489.1 GI:159220
 KEYWORDS anti-platelet protein.
 SOURCE Haementeria officialis cDNA to mRNA.

ORGANISM Haementeria officialis
 Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
 Rhynchobdellida; Glossiphoniidae; Haementeria.
 REFERENCE 1 (bases 1 to 563)
 AUTHORS Keller, P.M., Schultz, L.D., Condra, C., Karzewski, J. and Connolly, T.M.
 TITLE An inhibitor of collagen-stimulated platelet activation from the salivary glands of the Haementeria officialis leech. II. Cloning of the cDNA and expression
 JOURNAL J. Biol. Chem. 267 (10), 6899-6904 (1992)
 MEDLINE 92202247
 PUBMED 1551898
 FEATURES
 source location/Qualifiers
 1..563
 /organism="Haementeria officialis"
 /db_xref="taxon:6410"
 1..563
 /gene="LAPP"
 58..501
 /gene="LAPP"
 /codon_start=1
 /product="anti-platelet protein"
 /protein_id="AA029194.1"
 /db_xref="GI:159221"
 /translation="TMSPLPSLACSLIYAIAPISADPEDAGADDESEGEDTGSDE
 TPSTGGGSGDGNETITRAGNDCKSRGKLPNPLTKTEFTSVDECRKMCESAVE
 PSCYTIQINTETNECYRNNEGDVWSSLDQPNVQWHLHACSK"
 58..120
 /gene="LAPP"
 121..498
 /gene="LAPP"
 /product="anti-platelet protein"
 563
 /gene="LAPP"
 polyA-site
 BASE COUNT 190 a 114 c 127 g 132 t
 ORIGIN

Alignment Scores:

| Pred. No.: | 0.24 | Length: | 563 |
|------------------------|--------|---------------|-----|
| Score: | 85.50 | Matches: | 19 |
| Percent Similarity: | 51.79% | Conservative: | 10 |
| Best Local Similarity: | 33.93% | Mismatches: | 24 |
| Query Match: | 14.92% | Indels: | 3 |
| DB: | 3 | Gaps: | 1 |

US-09-936-737a-2 (1-103) x HEILAPP (1-563)

```

QY 4 GluAspCysTrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 23
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 250 GAAGATTTGCTGCTCTAAACGGCGCTGGCGAAGCTGCTGACATCTGTTGACGAGAACT 309

QY 24 LysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTyr----- 40
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 310 GAATTCACCGAGCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369

QY 41 CysTyrL1LevalPheGluAspThrValAsnLysGluCysTyrTyrAsn 56
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 370 TGTCTACATTCTGCAATACACAGAAACCAAGCAATCTATAGAAAT 417

```

RESULT 7
 LOCUS AL138835 95615 bp DNA linear PRI 13-APR-2001
 DEFINITION Human DNA sequence from clone RP11-554F20 on chromosome 9, complete sequence.
 ACCESSION AL138835
 VERSION AL138835.9 GI:13624930
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 95615)
 AUTHORS Lloyd, D.

TITLE Direct Submission
JOURNAL Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Apr 14, 2001 this sequence version replaced g1:13443296. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-554P20 is from the library RP11-547C13 is at 37658 in this sequence. The true left end of clone RP11-547C13 is at 37658 in this sequence. The true right end of clone RP11-547C13 is at 37658 in this sequence.

FEATURES

Source

location/Qualifiers

1..95615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-554P20"
/clone_lib="RP11-547C13"
114..351
/note="L2 repeat: matches 2341..2566 of consensus"
387..846
/note="L2 repeat: matches 21..466 of consensus"
1111..1407
/note="Alu repeat: matches 7..308 of consensus"
1864..2320
/note="match: STS: EM:HSG1D7"
1881..1925
/note="5 copies 9 mer tgctgtgtg 88% conserved"
2094..2286
/note="L2 repeat: matches 2009..2207 of consensus"
3322..3470
/note="MIR repeat: matches 61..207 of consensus"
3943..4199
/note="L2 repeat: matches 2423..2702 of consensus"
4307..4480
/note="MER58A repeat: matches 37..207 of consensus"
4517..4617
/note="MIR repeat: matches 123..232 of consensus"
5850..5932
/note="L2 repeat: matches 1988..2234 of consensus"
6248..6480
/note="MER96 repeat: matches 91..174 of consensus"
7258..7554
/note="L2 repeat: matches 1988..2234 of consensus"
8289..8499
/note="MIR repeat: matches 1..297 of consensus"
/note="MIR repeat: matches 8..237 of consensus"

repeat_region 8819..8901
/note="L2 repeat: matches 2335..2419 of consensus"
repeat_region 8927..8996
/note="L2 repeat: matches 2231..2299 of consensus"
repeat_region 9012..9531
/note="L2 repeat: matches 2180..2738 of consensus"
repeat_region 9586..9894
/note="L2 repeat: matches 2348..2747 of consensus"
repeat_region 10360..10424
/note="L2 repeat: matches 2646..2708 of consensus"
repeat_region 10421..10489
/note="MIR repeat: matches 48..118 of consensus"
repeat_region 11504..11584
/note="MIR repeat: matches 184..260 of consensus"
repeat_region 12099..12256
/note="MIR repeat: matches 16..190 of consensus"
repeat_region 13382..13680
/note="MIR repeat: matches 143..457 of consensus"
repeat_region 13681..14104
/note="MER1A repeat: matches 65..527 of consensus"
repeat_region 14129..14174
/note="MIR repeat: matches 411..495 of consensus"
repeat_region 15322..15616
/note="L2 repeat: matches 2656..2749 of consensus"
repeat_region 15613..15829
/note="L2 repeat: matches 2177..2418 of consensus"
repeat_region 17533..17662
/note="L2 repeat: matches 2613..2750 of consensus"
repeat_region 18242..18593
/note="MIR repeat: matches 1..365 of consensus"
repeat_region 19072..19266
/note="MER46C repeat: matches 118..319 of consensus"
repeat_region 19843..20184
/note="MIR repeat: matches 65..410 of consensus"
repeat_region 20734..20959
/note="MIR repeat: matches 10..250 of consensus"
repeat_region 21671..22060
/note="L2 repeat: matches 5777..6165 of consensus"
repeat_region 22327..22470
/note="MIR repeat: matches 13..163 of consensus"
repeat_region 22860..22957
/note="MIR repeat: matches 47..146 of consensus"
repeat_region 23642..23705
/note="MIR repeat: matches 68..132 of consensus"
repeat_region 24151..24345
/note="MIR repeat: matches 29..235 of consensus"
repeat_region 24365..24964
/note="L2 repeat: matches 5660..6286 of consensus"
repeat_region 24986..25254
/note="L2 repeat: matches 2459..2747 of consensus"
repeat_region 25934..26141
/note="MIR repeat: matches 1..253 of consensus"
repeat_region 26508..26640
/note="MIR repeat: matches 124..245 of consensus"
repeat_region 26655..26892
/note="MIR repeat: matches 6..262 of consensus"
repeat_region 26900..26992
/note="MIR repeat: matches 20..111 of consensus"
repeat_region 27024..27480
/note="L2 repeat: matches 1815..2289 of consensus"
repeat_region 27873..28179
/note="L2 repeat: matches 1..305 of consensus"
repeat_region 28301..28417
/note="MIR repeat: matches 1..122 of consensus"
repeat_region 28394..29077
/note="MIR repeat: matches 1..122 of consensus"
repeat_region 29079..29136
/note="L2 repeat: matches 2 mer at 70% conserved"
repeat_region 29188..29307
/note="L2 repeat: matches 2625..2748 of consensus"
repeat_region 29427..29487
/note="MIR repeat: matches 51..112 of consensus"

```

repeat_region 29591..29906
/note="AluY repeat: matches 1..310 of consensus"
repeat_region 30323..30712
/note="MT1B repeat: matches 7..390 of consensus"
repeat_region 33200..33272
/note="L2 repeat: matches 2626..2702 of consensus"
repeat_region 34778..34890
/note="MER5A repeat: matches 3..110 of consensus"
repeat_region 35583..35685
/note="MIR repeat: matches 81..180 of consensus"
repeat_region 37170..37243
/note="MIR repeat: matches 77..148 of consensus"
repeat_region 37581..38098
/note="L2 repeat: matches 2153..2750 of consensus"
repeat_region 38695..38938
/note="MIR repeat: matches 21..261 of consensus"
repeat_region 38938..39002
/note="MER5A repeat: matches 97..161 of consensus"
repeat_region 39559..39602
/note="L2 repeat: matches 2638..2720 of consensus"
repeat_region 40801..40832
/note="16 copies 2 mer ga 84% conserved"
repeat_region 40985..41325
/note="AluYB8 repeat: matches 1..318 of consensus"
repeat_region 41844..42158
/note="AluX8 repeat: matches 1..310 of consensus"
repeat_region 42382..42535
/note="MIR repeat: matches 95..252 of consensus"
repeat_region 42788..42921
/note="L2 repeat: matches 2555..2691 of consensus"
repeat_region 43117..43172
/note="28 copies 2 mer tt 73% conserved"
repeat_region 43604..43697
/note="L1M9 repeat: matches 5693..5786 of consensus"
repeat_region 43688..43818
/note="Tigref1 (Zomb1) repeat: matches 3..124 of consensus"
repeat_region 43860..44067
/note="MSTA repeat: matches 1..229 of consensus"
repeat_region 44088..44127
/note="20 copies 2 mer tc 90% conserved"
repeat_region 44129..44338
/note="MSTA repeat: matches 200..426 of consensus"

```

```

Alignment Scores:
Pred. No.: 197
Score: 83.00
Percent Similarity: 48.61%
Best Local Similarity: 36.11%
Query Match: 14.49%
DB: 9
Gaps: 2

```

```

US-09-936-737a-2 (1-103) x AL138835 (1-95615)
QY 4 GluspcysrtrpnrphetrAlaAsnArgLysrThrasphesplysSerPhelys 23
DB 95304 GAGCAGGTAGGCGGTAAAGCCGACGATCTCTATATGATTTTAAAAAGCA----- 95357
QY 24 LysSerSerAspLysAspGluCysLysLysThrCysPhelysThrGluTyr----- 40
DB 95358 ---GCCCTCAGAGATGCTGAGAGTTTGGATACGTTGTTGTAAGAAAGAGATTAAAGA 95414
QY 41 CystTtLleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValAspGly 60
DB 95415 TGTACACAAATTTTGGCCCTGGAACGTAAGAAAGAGATTATCATGTGCTGAAGAGG 95474
QY 61 GluGluLeuAspGluLysLysPheValValAspGlu 72
DB 95475 GAGCAGCTGTAAAGAGAGTGTGCTGAGAGCA 95510

```

```

RESULT 8
AC101018 149109 bp DNA linear HTG 21-AUG-2002
LOCUS Rattus norvegicus clone RP32-328P7, *** SEQUENCING IN PROGRESS ***,
DEFINITION

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

9 unordered pieces.
AC101018
AC101018.2 GI:22381320
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPERIN.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

1 (bases 1 to 149109)
Birren,B., Nusbaum,C. and Lander,E.
Rattus norvegicus, clone RP32-328P7
Unpublished
2 (bases 1 to 149109)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barre,N., Bastien,V., Boguslavsky,L., Bouknight,B.,
Brown,A., Camarata,V., Campolano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhahng,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Riedack,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talmas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE JOURNAL REFERENCE AUTHORS

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 149109)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barre,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouknight,B.,
Camarata,V., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunhahng,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talmas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE JOURNAL COMMENT

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L14794
Center clone name: 328_P7

| | | | | | |
|---|--|---------|---------|------------------------------|-------------------------------|
| | * | 65713 | 65812: | gap of unknown length | |
| | * | 65813 | 69339: | contig of 3527 bp in length | |
| | * | 69340 | 69438: | gap of unknown length | |
| | * | 69440 | 72570: | contig of 3131 bp in length | |
| | * | 72571 | 72670: | gap of unknown length | |
| | * | 72671 | 76898: | contig of 4228 bp in length | |
| | * | 76899 | 76998: | gap of unknown length | |
| | * | 76999 | 80760: | contig of 3762 bp in length | |
| | * | 80761 | 80860: | gap of unknown length | |
| | * | 80861 | 86313: | contig of 5453 bp in length | |
| | * | 86314 | 86413: | gap of unknown length | |
| | * | 92232 | 92231: | contig of 5818 bp in length | |
| | * | 92332 | 99692: | contig of 7361 bp in length | |
| | * | 99693 | 99792: | gap of unknown length | |
| | * | 99793 | 106178: | gap of 6366 bp in length | |
| | * | 106179 | 106278: | gap of unknown length | |
| | * | 106279 | 114304: | contig of 8026 bp in length | |
| | * | 114305 | 114404: | gap of unknown length | |
| | * | 114405 | 122112: | contig of 7708 bp in length | |
| | * | 122113 | 122212: | gap of unknown length | |
| | * | 122643 | 129742: | contig of 7430 bp in length | |
| | * | 129743 | 138684: | gap of unknown length | |
| | * | 138685 | 138784: | contig of 8942 bp in length | |
| | * | 138785 | 149662: | gap of unknown length | |
| | * | 149663 | 149762: | contig of 10878 bp in length | |
| | * | 149763 | 159492: | gap of unknown length | |
| | * | | | contig of 9730 bp in length. | |
| FEATURES | | | | | |
| SOURCE | | | | | |
| | | | | | Location/Qualifiers |
| | | | | | /organism="Rattus norvegicus" |
| | | | | | /db_xref="taxon:10116" |
| | | | | | /clone="CH230-286017" |
| BASE COUNT | 39083 | a | 36447 | c | 36055 g 40464 t 7443 others |
| ORIGIN | | | | | |
| Alignment Scores: | | | | | |
| Pred. No.: | | | | | 360 |
| Score: | | | | | 83.00 Length: 159492 |
| Percent Similarity: | | | | | 50.00% Matches: 20 |
| Best Local Similarity: | | | | | 30.30% Mismatches: 13 |
| Query Match: | | | | | 14.49% Indels: 31 |
| DB: | | | | | Gaps: 2 Gaps: 2 |
| US-09-936-f737A-2 (1-103) x AC115173 (1-159492) | | | | | |
| OY 39 | glutTTCysTrIleValphegIuasPhrYValasnlYsgIucSYrrTYrAsnValval | 58 | | | |
| Db 110192 | GACTCATGTAATGCAGCAGAGAACAAGAGGTGGAGCGACGACATCTACTATGACATTAC | 110133 | | | |
| OY 59 | AspgIcGluglUeuauspInclnuIsySpheValValAspgIuaSn---PherIngluaSn | 77 | | | |
| Db 110132 ---TATTACAAATACGAAATGAATAAACATCTACGAAGATGAAAACATCTACTATGANTAC | | 110076 | | | |
| OY 78 | TyrLeuThrAsPCysglueIlylysAspaIagIysanaIaaIagIyThrgIYaSpgluSer | 97 | | | |
| Db 110075 | TATGATGTTGACAGATGACCAATGACATATTGACATATTCAGATATACGACAGCAGCATGAT | 110016 | | | |
| OY 98 | AspgIuValAspgIuaSp 103 | | | | |
| Db 110015 | GACGATGTCGACGACGAC 109998 | | | | |
| RESULT 10 | | | | | |
| SFU15038 | SFU15038 | 2255 bp | mRNA | linear | INV 06-JAN-1995 |
| LOCUS | Spodoptera frugiperda | | | | |
| DEFINITION | immutunophilin FKBP46 mRNA, complete cds. | | | | |
| ACCESSION | U15038 | | | | |
| VERSION | U15038.1 GI:595844 | | | | |
| KEYWORDS | FK506 binding protein; immunophilin; peptidylprolyl cis-trans isomerase. | | | | |
| SOURCE | Spodoptera frugiperda | | | | |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | | |

```

55 rasnvalvalaspcylglutluleuaspclngclutlusphevalvalaspcgluasnpneth 75
:::|||||:::|||||:::|||||:::|||||
QY

```

| SOURCE | ORGANISM |
|------------------------|-----------------------|
| Plasmodium falciparum. | Plasmodium falciparum |

Web site: <http://www.hgsc.bcm.tmc.edu/>

Db 37552 ---CTGGAGAGACCAAGACGACA-----TATGCT----- 375526

Qy 47 AsptHrValAsnLysGluCysTyrTyrAsnValAlaSpGly 60
 Db 37525 GATGTTTGAATTAAGAAATGCTTTTATATCAATCAAGGA 37484

RESULT 14
 AC108277_2/c
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AC108277 Accession AC108277
 Fragment Name Begin End
 AC108277_0 1 110000
 AC108277_1 100001 210000
 AC108277_2 200001 310000
 AC108277_3 300001 375958
 Continuation (3 of 4) of AC108277 from base 200001 (AC108277 Rattus norvegicus clone CH2)

Alignment Scores:
 Pred. No.: 613 Length: 110000
 Score: 79.50 Matches: 21
 Percent Similarity: 43.488 Conservative: 9
 Best Local Similarity: 30.438 Mismatches: 16
 Query Match: 13.878 Indels: 23
 DB: 2 Gaps: 3

US-09-936-737a-2 (1-103) x AC108277_2 (1-110000)

Qy 3 ArgGluAspCysTyrPheTyrAlaAsn----- 12
 Db 24766 AGAGGAGAGAAATGGCTATTGCTCATTCGCTTCACAGTTTACTGGGTCAGATG 24707

Qy 13 -----ArgLysTyrThrAspPheAspLysSerPheLys--- 23
 Db 24706 CAGCCCTGTGTGTTCCAGAGATCACCCGCTACTCTACTATGACAACTGAAGTTG 24647

Qy 24 ---LysSerSerAspLeu-----AspGluCysLysLysThrCysPheLysThrGlu 39
 Db 24646 ACTCGCTCAATGAGCTAAAGTATAGAGACTCTTGTCCAAAGCCCTCTGCTAAACCTCT 24587

Qy 40 TYTCYSTYRLLIleValPheGluAspThr 48
 Db 24586 TACTGCTTCTGTGTAATTCAGCAACC 24560

RESULT 15
 AC103570/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-10W24, *** SEQUENCING IN PROGRESS
 AC103570
 VERSION
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 232370)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbora,J., Benton,J., Blincoe,K., Blankenburg,K., Bonaldi,D.,
 Bouck,J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Earhart,K.J., Drepper,H., Dugan-Rocha,S., Durbin,K.J.,
 Falls,T., Ferraguto,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsl,F., Howard,S., Huber,J., Huik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louseged,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,N., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
 Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,J., Rolfe,M., Ruiz,S., Savary,G.,
 Scheraga,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmali,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 Submitted (29-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 232370)
 Worley,K.C.
 Direct Submission
 Submitted (12-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced gl:20467456.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc.help@bcm.tmc.edu
 ----- Project Information
 Center project name: GDR0
 Center clone name: CH230-10W24
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 196541 bases at least Q40
 Consensus quality: 199662 bases at least Q30
 Consensus quality: 201786 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_diff_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 38 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1403: contig of 1403 bp in length
 * 1404 1503: gap of unknown length
 * 1504 2753: contig of 1250 bp in length
 * 2754 2853: gap of unknown length
 * 2854 4531: gap of unknown length
 * 4532 4631: contig of 1678 bp in length
 * 4632 4632: gap of unknown length
 * 4632 5823: contig of 1192 bp in length
 * 5823 5824: gap of unknown length

| Category | Count | Percentage |
|----------|-------|------------|
| 60783 a | 51173 | 84.03% |
| c | 49242 | 80.88% |
| g | 60940 | 100.00% |
| t | 10232 | 16.73% |
| others | 10232 | 16.73% |

Downloaded from <http://ajph.org/> on November 10, 2014

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:32:15 ; Search time 229 Seconds

(without alignments)
1012.908 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573

Sequence: 1 EEREDCWTFYANKKYTFDK.....GKDAGNAGTGDESDVED 103

Scoring table:

BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO-SPool/US0996737/runat_14022003.100732.2281/app.query.fasta.1.263
-DB=N_Geneseq_101002 -OFMT=fastap -SUPFIX=ring -MINMATCH=0.1 -LOOPTCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0996737 -GCN=1.118 -etunat_14022003.100732.2281 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -NAIT -LONGLOG -DEV TIMEOUT=120
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database:

N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 573 | 100.0 | 375 | 21 | AAA28180 | Platelet binding 1 |
| 2 | 81.5 | 14.2 | 2284 | 20 | AAO4441 | S. frugiperda immu |
| 3 | 80.5 | 14.0 | 4441 | 13 | AAQ23682 | leech antiplatelet |
| 4 | 77 | 13.4 | 1107 | 23 | ABU26101 | Drosophila melanog |
| 5 | 76.5 | 13.4 | 587 | 24 | ABQ47468 | Oligonucleotide fo |
| 6 | 76.5 | 13.4 | 587 | 24 | ABQ47469 | Oligonucleotide fo |
| 7 | 76 | 13.3 | 2496 | 20 | AAK20297 | Borrelia burgdorfe |
| 8 | 75.5 | 13.2 | 1399 | 16 | AAQ29969 | Human interleukin-1 |
| 9 | 75.5 | 13.2 | 1402 | 20 | AAQ31554 | Human full-length |
| 10 | 75.5 | 13.2 | 1402 | 20 | AAQ32645 | Human full-length |
| 11 | 75.5 | 13.2 | 1402 | 21 | AAQ2843 | Mouse caspase-12 |
| 12 | 75.5 | 13.2 | 4453 | 23 | ABU10668 | Drosophila melanog |
| 13 | 75 | 13.1 | 2149 | 22 | AAK94875 | Drosophila melanog |
| 14 | 74.5 | 13.0 | 2562 | 23 | ABV25467 | Human full-length |
| 15 | 74.5 | 12.8 | 1116 | 22 | AAK84453 | Human prostate exp |
| 16 | 73.5 | 12.8 | 3465 | 23 | AAK84456 | Human immune/haema |
| 17 | 73.5 | 12.8 | 3465 | 23 | ABU20199 | Drosophila melanog |
| 18 | 73.5 | 12.8 | 3465 | 23 | ABU20199 | Drosophila melanog |
| 19 | 72 | 12.6 | 27884 | 22 | AAK77781 | Human immune/haema |
| 20 | 71.5 | 12.5 | 762 | 24 | ABN70859 | Streptococcus poly |
| 21 | 71.5 | 12.5 | 789 | 24 | ABN68798 | Streptococcus poly |
| 22 | 71.5 | 12.5 | 797 | 24 | ABQ43620 | Oligonucleotide fo |
| 23 | 71.5 | 12.5 | 797 | 24 | ABQ43621 | Oligonucleotide fo |
| 24 | 71.5 | 12.5 | 963 | 21 | AAV70197 | Human immune/haema |
| 25 | 71.5 | 12.5 | 1092 | 23 | AAV58809 | Human immune/haema |
| 26 | 71.5 | 12.5 | 1106 | 21 | AAQ37864 | Human immune/haema |
| 27 | 71.5 | 12.5 | 1261 | 24 | ABO68909 | Human immune/haema |
| 28 | 71.5 | 12.5 | 1495 | 24 | ABO70444 | Human immune/haema |
| 29 | 71.5 | 12.5 | 1826 | 22 | AAH17523 | Human immune/haema |
| 30 | 71.5 | 12.5 | 21556 | 24 | ABN7157 | Streptococcus poly |
| 31 | 71 | 12.4 | 2299 | 23 | ABU22190 | Drosophila melanog |
| 32 | 71 | 12.4 | 9842 | 20 | AAK20259 | Borrelia burgdorfe |
| 33 | 70.5 | 12.3 | 572 | 22 | AAH1808 | Human differential |
| 34 | 70.5 | 12.3 | 661 | 24 | ABU2641 | Human differential |
| 35 | 70.5 | 12.3 | 1187 | 23 | ABU20275 | Human differential |
| 36 | 70.5 | 12.3 | 1370 | 23 | ABV24960 | Human prostate exp |
| 37 | 70.5 | 12.3 | 1370 | 23 | ABV25121 | Human prostate exp |
| 38 | 70 | 12.2 | 540 | 20 | AAV88360 | Nucleotide sequenc |
| 39 | 70 | 12.2 | 2010 | 24 | ABU56367 | Human prostate exp |
| 40 | 70 | 12.2 | 2010 | 24 | ABU56367 | Human prostate exp |
| 41 | 70 | 12.2 | 2095 | 23 | ABU2989 | Human prostate exp |
| 42 | 70 | 12.2 | 3504 | 23 | AAV3788 | Helicobacter pylori |
| 43 | 70 | 12.2 | 4994 | 23 | ABU2989 | Helicobacter pylori |
| 44 | 69.5 | 12.1 | 344 | 22 | AAV39459 | Drosophila melanog |
| 45 | 69.5 | 12.1 | 344 | 22 | AAK88317 | Human digestive sy |

ALIGNMENTS

RESULT 1
AAA28180 standard; cDNA: 375 BP.

29-JAN-2001 (first entry)

Platelet binding inhibitor protein Saratin encoding cDNA sequence.

Saratin: medicinal leech; thromboembolic disease; intraocular lens;
collagen-dependent platelet adhesion inhibition;
posterior capsule opacification; ss.

Hirudo medicinalis.
Key Location/Qualifiers
CDS 64..375

SAWE


```

FT      /*tag= a
FT      /product= "Saratin"
FT      /note= "Platelet adhesion inhibitor protein"
XX
XX      WO200056885-A1.
XX
XX      28-SEP-2000.
XX
XX      10-MAR-2000; 2000WO-EP02117.
XX
XX      18-MAR-1999; 99EP-0105530.
XX
XX      12-MAY-1999; 99EP-0109503.
XX
XX      (MERE ) MERCK PATENT GMBH.
XX
XX      Stiltmutter W, Guessow D, Hofmann U, Hemberger J, Fotev Z;
XX      Scheuble B;
XX      WPI: 2000-611629/58.
XX      P-PSDB; AAY94746.
XX
XX      New Saratin polypeptide and gene isolated from Hirudo medicinalis for
XX      blocking platelet adhesion, especially useful for treating or
XX      preventing thrombotic diseases, or for manufacturing a medicament for
XX      thromboembolic diseases
XX
XX      Claim 4; Page 40; 46pp; English.
XX
XX      Saratin is a protein isolated from the saliva of the medicinal leech
XX      Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
XX      platelet adhesion. The invention includes polynucleotide sequences, a
XX      encoding Saratin, an expression vector comprising the DNA sequence, a
XX      host cell transformed with the expression vector, antibodies
XX      immunospecific for Saratin, and methods for identifying thromboembolic processes,
XX      or antagonists. Saratin is useful for treating thromboembolic diseases,
XX      and for manufacturing a medicament for treating thromboembolic diseases.
XX      It is useful for preventing thrombotic diseases. Saratin is also useful
XX      for coating artificial surfaces, since use of Saratin renders them
XX      non-adhesive for cells and prevents the activation of cells. It may also
XX      be used for coating natural collagen surfaces. Furthermore, Saratin is
XX      useful for modifying intraocular lenses in order to lessen the
XX      thrombogenicity of the lens material, for contacting the lens surface, or
XX      for covalent crosslinking to modify the lens material. The lens material
XX      is used for refractive anterior or posterior chamber ocular implants,
XX      which may be implanted into the eye. This new type of coating avoids
XX      problems contributed by stimulated cell growth. In combination with other
XX      medicaments that are for instance conferring cell death, Saratin coating
XX      helps to completely overcome posterior capsule opacification. The
XX      antibody immunospecific for Saratin, as well as Saratin itself, are
XX      useful for measuring samples derived from host cell cultures or from a
XX      treated subject. The present sequence represents cDNA encoding the
XX      Saratin protein.
XX
XX      Sequence 375 BP; 123 A; 62 C; 85 G; 105 T; 0 other:
XX
XX      Alignment Scores:
XX      Pred. No.: 7 28e-64 Length: 375
XX      Score: 573.00 Matches: 103
XX      Percent Similarity: 100.008 Conservative: 0
XX      Best Local Similarity: 100.008 Mismatches: 0
XX      Query Match: 100.008 Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-09-936-737a-2 (1-103) x AAA28180 (1-375)
XX
QY      1 GIUGIUAAGGUAAPCYSTIRPThrPheTyrAlaAsnArgLysTyrThrAspPheAspLys 20
DB      64 GAAGAACGTAAGATTGTTGACGTTTACCGCAACGAAATATATACAGACTTCGATATAA 123
QY      21 SerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTyr 40
DB      124 TCTTTTAAAGAAAGTCTTGATCTTGACGAATGCAAAAACATGTTTCAGACGAGATAC 163

```

```

QY      41 CysTyrTlleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValAspLys 60
DB      184 TGCATCATCGTTTGTGAACACGTCACACAGAAATGTTACTACAAATGCTGATGCT 243
QY      61 GluGluLeuAspGlnGluLysPheValValAspGluAsnPheThrGluAsnTyrLeuThr 80
DB      244 GAAGAGTTAGACCAAGCAAAAATTTGTTGTCAGCAAGAAACCTCACGAAATATTTTGACA 303
QY      81 AspCysGluGluLysAspAlaGlyAsnAlaAlaGlyThrGlyAspGluSerAspGluVal 100
DB      304 GACTGCGAGGGGTAAAGATGACGATATGCGCAGGTACAGGTGACAGTCAATGAGTT 363
QY      101 AspGluAsp 103
DB      364 GATGAAAGAT 372

RESULT 2
AAK04441
AAK04441 standard; cDNA; 2284 BP.
AAK04441;
XX      27-APR-1999 (first entry)
XX
XX      S. frugiperda immunophilin FKBP46 cDNA.
XX
XX      Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;
XX      transplant; tissue graft; ds.
XX
XX      Spodoptera frugiperda.
XX
XX      Key Location/Qualifiers
XX      CDS 160..1398
XX      FT /*tag= a
XX      FT /product= "immunophilin FKBP46"
XX      FT polyA_signal 1764..1769
XX      FT polyA_signal 1850..1855
XX      FT polyA_signal /*tag= b
XX      FT polyA_signal 1868..1873
XX      FT polyA_signal /*tag= c
XX      FT polyA_signal 2052..2057
XX      FT polyA_signal /*tag= d
XX      FT polyA_signal 2237..2242
XX      FT polyA_signal /*tag= e
XX      FT polyA_signal /*tag= f
XX
XX      US5861498-A.
XX      19-JAN-1999.
XX
XX      31-OCT-1996; 96US-0741134.
XX
XX      01-NOV-1995; 95US-0007163.
XX      31-OCT-1996; 96US-0741134.
XX
XX      (UWJE-) UNIV JEFFERSON THOMAS.
XX
XX      Alnemri ES, Fernandes-Alnemri T, Litwack G;
XX      WPI: 1999-130433/11.
XX      P-PSDB; AAW68010.
XX
XX      Isolated immunophilin FKBP46 nucleic acids - useful for developing
XX      products for the study and identification of immunosuppressive
XX      agents for treating e.g. transplant and tissue graft patients
XX
XX      Claim 2; Fig 1; 21pp; English.
XX
XX      This sequence represents the cDNA encoding the immunophilin FKBP46 from
XX      the moth Spodoptera frugiperda. FKBP46 is a novel insect cell nuclear
XX      immunophilin, which can be used to study and identify additional
XX      immunosuppressive drugs that bind to it. Such immunosuppressive agents
XX      can be used in treating transplant and tissue graft patients.

```

100

XX Claim 1: SEQ ID NO 29776; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB161175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 1107 BP; 284 A; 313 C; 308 G; 202 T; 0 other;

Alignment Scores:

| Pred. No.: | 3.48 | Length: | 1107 |
|------------------------|--------|---------------|------|
| Score: | 77.00 | Matches: | 28 |
| Percent Similarity: | 41.88% | Conservative: | 21 |
| Best Local Similarity: | 23.93% | Mismatches: | 34 |
| Query Match: | 13.44% | Indels: | 34 |
| DB: | 23 | Gaps: | 6 |

US-09-936-737a-2 (1-103) x AB126101 (1-1107)

QY 14 LysTyrThrAspPheAspLysSer-----PheLys 23
 DB 265 GAGTACACAGATGCGGAGAGAGCGATCCGCCACAGTTCCTACCCAGTTCCG 324
 QY 24 LysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTyrCysTyrile 43
 DB 325 TCCGCCCTGATGCTAAAGGCTAAAGGCAAAAGCA-----TTGAAATATGCAAAATG 375
 QY 44 ValPheGlu-----AspThrValAsnLysGluCysTyrTyrAsnValValAspLys 60
 DB 376 ATCTCGAATACGAGCCCGACGCCCAAGGCAAGTTCATCCCTCATC----- 429
 QY 61 GluGluLeuAspGlnGluLysPheValVal-----AspGluAsnPheThr 75
 DB 430 -----CTGGACAAAGTTGGGGCAGTGGCCACATCCAGCAGCAGATGAGACTACAAAT 483
 QY 76 GluAsn-----TyrLeuThrAspCysGluGlyLysAsp 86
 DB 484 AATATCTTCAATCCGCCGACTTGGCTGATGCTGCATGCGATGCGAGTGGAGCTGACGTG 543
 QY 87 AlaGlyAsnAlaAlaGlyThrGlyAspGluSerAspGluValAspGluAsp 103
 DB 544 GACGGGGATGAGCAGCAGAGATGCTGACGAGATGGGAGTGTGATGGCGAT 594

RESULT 5
 ABQ47468
 ID ABQ47468 standard; DNA; 587 BP.
 XX
 AC ABQ47468;
 XX
 DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34059.
 XX
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP10074.

XX
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPig-) EPiGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Gnetig D;
 XX
 DR WPI; 2002-371829/40.
 XX
 PR Determining the degree of cytosine methylation in genomic DNA, useful
 PR for diagnosis and prognosis, comprises selective hybridization of
 PR amplicons from chemically treated DNA
 XX
 PS Claim 12: 56bp + Sequence Listing; 56bp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridization to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridized to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

SQ Sequence 587 BP; 97 A; 81 C; 214 G; 195 T; 0 other;

| Alignment Scores: | 1.78 | Length: | 587 |
|------------------------|--------|---------------|-----|
| Pred. No.: | 76.50 | Matches: | 26 |
| Score: | 48.65% | Conservative: | 10 |
| Percent Similarity: | 35.14% | Mismatches: | 25 |
| Best Local Similarity: | 13.35% | Indels: | 13 |
| Query Match: | 24 | Gaps: | 5 |

US-09-936-737a-2 (1-103) x ABQ47468 (1-587)

QY 38 ThrGluTyrCysTyrTrlValPheGluAsp-----ThrValAsnLysGluCysTyrTyr 55
 DB 257 ACGGACGATGTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 316
 QY 56 Asn-----ValValAspGlyGluGluLeuAspGlnGluLysPheValValAspGluAsn 73
 DB 317 AACGATTGCTGTGCGACGGT-----TTCCAGCGGTTTATTTGCG---GAGTCG 361
 QY 74 PheThrGluAsnTyrLeuThrAspCysGluGlyLysAspAlaGlyAsnAlaGlyThr 93
 DB 362 GTTAAAGGAAATATGATGCTTGTGTTGGCGTGCAGTGTGATGATGATGATGATGAT 421
 QY 94 GlyAspGluSer-----AspGluValAspGluAsp 103
 DB 422 AGCGATTGACAGGGGGGTTGGACGATTATTACGATGATGATGATGATGATGAT 463

RESULT 6
 ABQ47469/C
 ID ABQ47469 standard; DNA; 587 BP.
 XX
 AC ABQ47469;
 XX
 DT 12-JUL-2002 (first entry)

```

Db      226 GTTAAAGAGAAATATGTTGCTTTCGTTCGGCGTCGACATTCGGGATACCGATTCCGTACT 16
QY      94 GLYAspGlnSer-----AspGluValAspGluAsp 103
      |||  ::|  |||  ||| ||| ||| ||| |||
Db      166 AGCGATTGACGCGGGGTTTGACGACGATTATTACGATGAGGAT 125
RESULT 7
AAAX20297/c
ID      AAAX20297 standard; DNA; 2496 BP.
XX
XX      AAAX20297;
XX
XX      04-MAY-1999 (first entry);
DE      Borrelia burgdorferi polynucleotide sequence #50.
XX
XX
XX      Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW      epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX      infection; diagnosis; characterisation; detection; ds.
XX
XX      Borrelia burgdorferi.
XX
XX      MO9858943-X1.
XX
XX      30-DEC-1998.
XX
XX      18-JUN-1998; 98MO-US12764.
XX
XX      03-SEP-1997; 97US--0057483.
XX      PR      20-JUN-1997; 97US--0050359.
XX      PR      22-JUL-1997; 97US--0053344.
XX      PR      22-JUL-1997; 97US--0053377.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI      White OR;
DR
DR      WPI; 1999-081217/07.
XX
XX      New isolated Borrelia burgdorferi nucleic acids - used to develop
PT      products for the detection, diagnosis, characterisation, prevention
PT      and therapy of infections, particularly Lyme disease
PS
PS      Claim 1; Page 1026-1028; 1128pp; English.
XX
XX      AAAX20248 to AAAX20402 represent polynucleotide sequences isolated from
CC      Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC      the detection, diagnosis, characterisation, prevention and therapy of
CC      Bb infections, e.g. Lyme disease. They can also be used for the
CC      production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC      to a family of motile, spiral-shaped bacteria called Spirochetes.
CC      Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC      endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC      Lyme disease.
XX
XX
SQ      Sequence 2496 BP; 871 A; 348 C; 235 G; 1029 T; 13 other;
Alignment Scores:
Pred. No.: 13.3 length: 2496
Score: 76.00 Matches: 18
Percent Similarity: 48.61% Conservative: 17
Best Local Similarity: 25.00% Mismatches: 25
Query Match: 13.26% Indels: 12
DB: Gaps: 3
US-09-936-737A-2 (1-103) x AAAX20297 (1-2496)
QY      10 TYRAlaAsnArgLysTYRThrAspPheAspLysSerPheLysLysSerSerAspLeuAsp 29
      |||  ::|  |||  ||| ||| ||| ||| |||
Db      344 TATGACGAGAAAAAAGATATTATTGATTGAT-----TTTCTTAATAAATTAAAGAAAAAGAA 291

```



```

Query Match:      13.18%      Indels:      19
DB:               17         Gaps:          6
US-09-936-737A-2 (1-103) x AA331554 (1-1399)

Qy  44 ValhEgIuAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
    |||||:|||||  |||  |||  |||||:|||||  |||||:|||||  |||||:|||||
Db  120 GTTTTGTATGACCTGGTGAGAGAG-----AACTTTTAAATGAGATGAGTTG 167
Qy  64 -----AspGluGluLysPheValValAsp-----GluAsnDheThrgLusn 77
    :|||:|||||  |||||:|||||  |||||  |||||  |||||
Db  168 CTCAAAATAGGGGAAAGTCGAGTTTCATCTCGAACAGCTGAGATCTGTTGAGAAC 227
Qy  78 TyrLeu-----ThraSpCysGluGlyLys---AspAlaGlyAsnAlaIaGlyThrgly 94
    :|||:|||||  |||||  |||||  |||||:|||||  |||||:|||||
Db  228 TTCTTAGAGAAACAGACATGCGAGAAAATTTTGCTGCGCACATTCCTCATTCACAG 287
Qy  95 AspGlu-----SerAspGluValAsp 101
    :|||:|||||  |||||  |||||  |||||
Db  288 GAACAGCTGAGTTTACATTTTCTATATGATGAGATGAT 326

RESULT 10
AA332645
ID  AA332645 standard; cDNA: 1402 BP.
XX
AC  AA332645;
XX
DT  21-JAN-2000 (first entry)
XX
DE  Murine caspase-12 splice variant, caspase-12L cDNA.
XX
KW  Caspase; splice variant; truncated; programmed cell death; apoptosis;
KW  regulation; proteolytic cascade; malignant condition; cancer;
KW  premalignant condition; solid tumour; lymphoma;
KW  chronic lymphocytic leukaemia; prostatic hypertrophy;
KW  preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
KW
XX
OS  Mus sp.
XX
FH  Key
FT  CDS
FT  1..1050
FT  Location/Qualifiers
FT  /*tag= a
FT  /product= "Murine caspase-12 splice variant,
FT  /product= caspase-12L"
FT  exon
FT  1..93
FT  /*tag= b
FT  /number= 1
FT  94..303
FT  /*tag= c
FT  /number= 2
FT  304..1402
FT  /*tag= d
FT  complement (918..943)
FT  /*tag= e
FT  /bound_molecly= "Degenerate upstream PCR primer AA332648"
FT  1281..1300
FT  /*tag= f
FT  /bound_molecly= "Degenerate downstream PCR primer
FT  AA332649"
XX
XX  WO9952925-A1.
XX
XX  21-OCT-1999.
XX
XX  14-APR-1999; 99MO-US08064.
XX
XX  16-APR-1998; 98US-0081962.
XX  (GEHO ) GEN HOSPITAL CORP.
XX
XX  Yuan J, Morishima N;
XX

```

```

DR  WPI: 1999-620369/53.
DR  P-PSDB: AA550109.
XX
PT  New nucleic acid encoding the short form of caspase-12, used e.g. for
PT  treating tumors
XX
PS  Example 1; Fig 1; 68pp; English.
XX
CC  This sequence represents a cDNA encoding murine caspase-12L, a splice
CC  variant of caspase-12. Caspases are a family of proteins involved in
CC  the regulation of apoptosis and are synthesised as proforms which are
CC  activated via cleavage after specific Asp residues. Mammalian cells
CC  express several caspases, and it is thought that these act in a
CC  proteolytic cascade to cause programmed cell death. Nucleic acids
CC  encoding caspase-12S (AA332644) or truncated forms of caspase-12L
CC  (AA332646, AA332647) are used for production, recombinantly or in vivo,
CC  of caspase-12 polypeptides which induce programmed cell death. This is
CC  particularly useful for treating (pre)malignant conditions (e.g., solid
CC  tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
CC  hypertrophy, preneoplastic liver foci and resistance to chemotherapy), or
CC  autoimmune diseases. The caspase-12 proteins can also be used to raise
CC  specific antibodies (for example, to determine gene expression and to
CC  screen expression libraries) or as molecular weight markers. Fragments
CC  of caspase-12-encoding nucleic acids can be used as probes to isolate
CC  the caspase-12 gene (and its allelic variants); in fluorescent in situ
CC  hybridisation for chromosomal location of the caspase-12 gene, and for
CC  Northern blotting to determine caspase-12 mRNA expression in tissues.
XX
SQ  Sequence 1402 BP; 442 A; 270 C; 315 G; 375 T; 0 other;

Alignment Scores:
Pred. No.:      7.31      Length:      1402
Score:          75.50      Matches:      26
Percent Similarity: 54.79%      Conservative: 14
Best Local Similarity: 35.62%      Mismatches: 14
Query Match:    20        Indels:      19
DB:             20        Gaps:       6

US-09-936-737A-2 (1-103) x AA332645 (1-1402)

Qy  44 ValhEgIuAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
    |||||:|||||  |||  |||  |||||:|||||  |||||:|||||  |||||:|||||
Db  120 GTTTTGTATGACCTGGTGAGAG-----AACTTTTAAATGAGATGAGTTA 167
Qy  64 -----AspGluGluLysPheValValAsp-----GluAsnDheThrgLusn 77
    :|||:|||||  |||||:|||||  |||||  |||||  |||||
Db  168 CTCAAAATAGGGGAAAGTCGAGTTTCATCTCGAACAGCTGAGATCTGTTGAGAAC 227
Qy  78 TyrLeu-----ThraSpCysGluGlyLys---AspAlaGlyAsnAlaIaGlyThrgly 94
    :|||:|||||  |||||  |||||  |||||:|||||  |||||:|||||
Db  228 TTCTTAGAGAAACAGACATGCGAGAAAATTTTGCTGCGCACATTCCTCATTCACAG 287
Qy  95 AspGlu-----SerAspGluValAsp 101
    :|||:|||||  |||||  |||||  |||||
Db  288 GAACAGCTGAGTTTACATTTTCTATATGATGAGATGAT 326

RESULT 11
ID  AA332643 standard; DNA: 1402 BP.
XX
XX  AA332643;
XX
XX  09-FEB-2001 (first entry)
XX
XX  Mouse Ice-4 coding sequence.
XX
XX  ced-3; virally induced cell death; apoptosis; gene therapy; neural;
XX  muscular degenerative disease; myocardial infarction; stroke; aging;
XX  Interleukin-1beta converting enzyme; ICE; mouse;
XX  Ice-ced 3 homologue; Ich; ss.
XX
OS  Mus sp.
XX

```

Key Location/Qualifiers
 CDS 45..1304
 /*tag= a
 /product= "Mouse Ice-4"

US6083735-A.
 04-JUL-2000.
 10-JUN-1994; 94US-0258287.
 24-JUN-1993; 93US-0080850.
 (GENO) GEN HOSPITAL CORP.
 Yuan J, Miura M;
 WPI; 2000-464343/40.
 P-PSDB; AAB14259.

New human Ich-1L and Ich-1S proteins for negative and positive regulation of programmed cell death and for developing therapeutic methods for diseases and conditions characterized by cell death, e.g. myocardial infarction or stroke
 Disclosure; Fig 16; 121pp; English.

The present sequence is the mouse Ice-4 protein coding sequence. The protein sequence of the present sequences was used in a sequence homology comparison with the protein sequences of human Ice-3 homolog (Ich-1) (AAB14253), murine Interleukin-1beta converting enzyme (mICE) (AAB14249), murine ICE2 (AAB14252) and C. elegans ced-3 (AAB14246). The present sequence is a member of a family of genes involved in programmed cell death (apoptosis). Ich-1 may play an important role in both the positive and negative regulation of apoptosis. The Ich gene may be used in gene therapy in disorders characterised by cell death e.g. neural and muscular degenerative diseases, myocardial infarction, stroke, virally induced cell death and aging.

Sequence 1402 BP; 442 A; 270 C; 315 G; 375 T; 0 other;

Alignment Scores:
 Pred. No.: 7.31 Length: 1402
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 13.18% Indels: 19
 Gaps: 6

US-09-936-737A-2 (1-103) x AAA72843 (1-1402)

OY 44 ValPheGluAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
 |||||
 DB 120 GTTTTGTAGACCTGCTGACAGC-----AATCTTTAAATGAGATGAGTGA 167
 |||||
 OY 64 -----AspGluGluLysPheValValAsp-----GluAsnPheThrGluAsn 77
 |||||
 DB 168 CTCGAAATATAGGGAAGCGAGCTTCATCTGTAACAGCTGAGATCTGTTGAGAAC 227
 |||||
 OY 78 TyrLeu-----ThrAspCysGluGlyLys-----AspAlaGlyAsnAlaIleGlyThrGly 94
 |||||
 DB 228 TTCTTAGAGAAACAGACAGATCTGCAAAATATTTCTGCGCCCATTTGCCAATTCACAG 287
 |||||
 OY 95 AspGlu-----SerAspGluValAsp 101
 |||||
 DB 288 GACAGCTGAGTTTACATTTTCTTAATCATGAGATGAT 326
 |||||

RESULT 12
 ABL10668
 ID ABL10668 standard; CDNA: 4453 BP.
 XX ABL10668;
 AC
 XX

26-MAR-2002 (first entry)
 Drosophila melanogaster expressed polynucleotide SEQ ID NO 26486.
 Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical; gene; ss.
 Drosophila melanogaster.
 WO200171042-A2.
 27-SEP-2001.
 23-MAR-2001; 2001WO-US09231.
 23-MAR-2000; 2000US-191637P.
 11-JUL-2000; 2000US-0614150.
 (PEKE) PE CORP NY.
 Venter JC, Adams M, Li PWD, Myers EW;
 WPI; 2001-656860/75.
 P-PSDB; ABB65655.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
 Claim 1; SEQ ID NO 26486; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB16175), expressed DNA sequences (AB16176-AB16175) and the encoded proteins (AB16176-AB16175).
 The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 4453 BP; 1232 A; 1020 C; 1060 G; 1141 T; 0 other;

Alignment Scores:
 Pred. No.: 32.4 Length: 4453
 Score: 75.50 Matches: 30
 Percent Similarity: 48.91% Conservative: 15
 Best Local Similarity: 32.61% Mismatches: 36
 Query Match: 13.18% Indels: 11
 Gaps: 6

US-09-936-737A-2 (1-103) x ABL10668 (1-4453)

OY 4 GluAspCysThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 23
 |||||
 DB 2124 GAGCAGGCCCGTACGCTCTGCGCAGCAGCAACACAA-----AAAAGTTTGAG 2174
 |||||
 OY 24 LysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluThrCys----- 41
 |||||
 DB 2175 GCTGCTGAGATCTTCGCGTTCCTGCAATATATGCTT-----CAAGTATGTAAATA 2228
 |||||
 OY 42 ---TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValValAspGly 60
 |||||
 DB 2229 GTTATATATGTA---GAATTTTCATTAATGTGAATGCTTCAGGAATTCGTGTAA 2285
 |||||
 OY 61 GluGluLeuAspGluGluLysPheValValAspGluAsnPheThrGluAsnTyrLeuThr 80
 |||||
 DB 2286 TTTGAGGCCAGCTGAGACCCGCTGCTCAGAGAGATGTTAAAGATAT---CTCAC 2342
 |||||
 OY 81 AspCysGluGlyLys-----AspAlaGlyAsnAlaIle 91
 |||||

Db 2343 GATAGTACTACAAATATCCGATTTAATGCCGCG 2378
 RESULT 13
 AAK94875
 ID AAK94875 standard; cDNA; 2149 BP.
 XX
 AC AAK94875;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 4066.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI30094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
 PI Wakematsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI: 2001-524255/58.
 DR P-PSDB; AAM93913.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation
 XX
 PS Claim 8; SEQ ID NO 4066; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones; 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SO Sequence 2149 BP; 659 A; 415 C; 531 G; 544 T; 0 other;

Alignment Scores:
 Pred. NO.: 14.7 Length: 2149
 Score: 75.00 Matches: 27
 Percent Similarity: 41.51% Conservative: 17
 Best Local Similarity: 25.47% Mismatches: 35
 Query Match: 13.09% Indels: 27
 DB: 22 Gaps: 4

US-09-936-737A-2 (1-103) x AAK94875 (1-2149)

QY 2 GUARGGLUASPCYSTTRPHTRPHEHYR-----AlaAsnArgIysTyrThrAspPhe 18
 Db :||||| ||||||| |||:| ||||| |||||
 Db 650 AAAAGAAGAACGCTGGTATTTTCCCTTGTCTGTAAGTTAAATATGTCATTTTC 709
 QY 19 AsplysSerPhe-----LysLysSerSer 26
 Db :||||| |||||
 Db 710 TCCCTATCATTTTGTATTAATTCCTGGAACATTTCTAGAGAAATTAACCTTCACCTCT 769
 QY 27 AspleuAspGluCysLysLysThrCysPheLysThrGluTyrCysTyrIle-ValPheG1 46

Db 770 TCCGTGACAGCTTTAATTGAAAAATTTTCCCAAGAGACTAT--TACTGTGATGCTC 826
 QY 46 uAspThrValAsnLysGluCysTyrTrpAsnValValAspGlyGluGluLeuAspGlnG1 66
 Db :||||| |||:| ||||| |||||
 Db 827 AGAGTCAGCAATCATGCTCTGATTTATGTTTCATTCACAAAGAGTGAAGATGAGCAACA 886
 QY 66 uLys-----PheValValAspGluAsnPheThrG1 76
 Db :||||| |||:| ||||| |||||
 Db 887 AAAAAGAACGAGTATTTGTGACAAAGGCTATTCGTGGTGGAGTCATCTTCAGTGG 946
 QY 76 uAsnTyrLeuThrAsp 81
 Db :||||| |||
 Db 947 TGAATGTTAGAAGAT 962

RESULT 14
 ABV25467
 ID ABV25467 standard; cDNA; 2562 BP.
 XX
 AC ABV25467;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 25458.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PI Schlegel R, Endege WO, Monahan JR;
 XX WPI: 2001-662795/76.
 DR

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 5054; 11750bp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
 (a) assessing whether a patient is afflicted with prostate cancer;
 (b) monitoring the progression of prostate cancer in a patient;
 (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
 (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
 (e) selecting a composition for inhibiting prostate cancer in a patient;
 (f) assessing the prostate cell carcinogenic potential of a compound;
 (g) determining whether prostate cancer has metastasized in a patient;
 (h) assessing the aggressiveness or indolence of prostate cancer in a patient;
 (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 2562 BP; 970 A; 365 C; 641 G; 576 T; 10 other;

Alignment Scores:
Pred. No.: 21.3 Length: 2562
Score: 74.50 Matches: 34
Percent Similarity: 45.76% Conservative: 20
Best Local Similarity: 28.81% Mismatches: 44
Query Match: 13.00% Indels: 20
DB: 23 Gaps: 7

US-09-936-737a-2 (1-103) x ABV25467 (1-2562)

QY 2 GLUARGGLAASPCYSTRPHRYR-----AlaasnArgLysrThrasp-Ph 18
Db 657 GAAAAAATGTTCTCAACATCTCAGACACTCTCTCCGAGAAAAACAAGACACTT 716
QY 18 eaaplySerPheLysLys-----SerSerAspleuaspGluCysLysLysThCysPh 36
Db 717 AGACTCAGCAGCTCTGAATCTGCAATCTCCAGAAATCGAGTGTCTTAAGACA----- 771
QY 36 elysrThglutyrCysrTyleVal---PheGluaspThrValasnLysGluCysrTy 55
Db 772 -AGAAAGAAATCAATCACTGCTCACTCAATAGACAGACAGACAGTATGA 830
QY 55 rasnValaspGlygluGluLeuaspGluLysPheValasp----- 71
Db 831 AAATCTCAACAGATGCTGAATGTGTGACAAAGATGCTGAGAAAGATTCAAAAGCGT 890
QY 72 -----GluasnPheThrGluasnTyrLeuThrAspCysGluGlyLysAs 86
Db 891 TACTGAATAGAGAGTCAATGAGAAATCTGAATGAATTAACAGTGT---GSTRAGAC 947
QY 86 PALAGLYASpAlaagLysrGlyaspGluSeraspGluValaspGluasp 103
Db 948 TTCAGGTGATGACGATGGAAGTGAAGATGATGAAGAGAG---GATGAAGAT 996

RESULT 15
AAK84453 standard; DNA; 1116 BP.
XX AAK84453;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39265.
XX
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosstatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0198974.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220965.
XX 26-JUL-2000; 2000US-0220964.
XX

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
50 Sequence 1116 BP; 315 A; 207 C; 208 G; 386 T; 0 other;

Alignment Scores:

| | | | |
|------------------------|--------|---------------|-----|
| Pred. No.: | 9.79 | Length: | 116 |
| Score: | 73.50 | Matches: | 22 |
| Percent Similarity: | 51.61% | Conservative: | 10 |
| Best Local Similarity: | 35.48% | Mismatches: | 21 |
| Query Match: | 12.83% | Indels: | 9 |
| OB: | 22 | Gaps: | 4 |

US-09-936-737A-2 (1-103) x AAK84453 (1-1116)

| | | | | |
|----|-----|--|--|-----|
| QY | 7 | trpItrpIhpe-- | TyrAlaIaasnArgLysTyrThrAspPheAspLysSerPheLysLysSer | 25 |
| | | | | |
| Db | 476 | TGAGTCGCTTAACACTGACAGAGCATTTT | TTTTTTTTTTTTTTTTTTTTTTTGGCT | 535 |
| | | | | |
| QY | 26 | SerAspIleu----- | AspGluCys--LysLysThrCysPheLysThrGluTyrCys | 41 |
| | | | | |
| Db | 536 | TCTGAGCGTGTTTTACAACTGACACCGTGTTCACACAGATGATTTTTCATAGCGTTTTGT | 595 | |
| | | | | |
| QY | 42 | TyrIleValPheGluAspThrValaLnLysGluCysTyrTyrAsnValaIaAspGlu | 61 | |
| | | | | |
| Db | 596 | TACTTAGTTCATCATGATGCC----- | TGTTATTTTTCCTGTGTAATTTAAAT | 643 |
| | | | | |
| QY | 62 | GluIeu | 63 | |
| | | | | |
| Db | 644 | GAGCTG | 649 | |
| | | | | |

Search completed: February 25, 2003, 02:38:20
Job time : 233 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:33:25 ; Search time 48 Seconds

(without alignments)
658.077 Million cell updates/sec

Title: US-09-936-737A-2

Perfect score: 573
Sequence: 1 EEREDCWFIFANRRKRYDFDK.....GKDGNAAGTDESDVEDD 103

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=framed+ .p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09936737/rnat_14022003.100733.2308/app_query.fasta.1.263
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=tni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human4.0.cdi
-LIST=43 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEXT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936737 @CGR.1.1.25 @unat.14022003.100733.2308 -NCPU=6 -ICPU=3
-NO.XLPHY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
4: /cgn2.6/prodata/1/ina/5B.COMB.seq.*
5: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
6: /cgn2.6/prodata/1/ina/5B.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 81.5 | 14.2 | 2255 | 2 | US-08-741-134-1 |
| 2 | 80.5 | 14.0 | 378 | 1 | US-08-044-547-4 |
| 3 | 80.5 | 14.0 | 441 | 1 | US-08-044-547-2 |
| 4 | 75.5 | 13.2 | 1402 | 3 | US-08-258-287B-56 |
| 5 | 75.5 | 13.2 | 1402 | 3 | US-08-368-704C-54 |
| 6 | 69.5 | 12.1 | 2023 | 3 | US-08-961-083-199 |
| 7 | 69.5 | 12.1 | 32768 | 4 | US-08-961-527-71 |
| 8 | 68 | 11.9 | 1540 | 4 | US-09-149-476-151 |
| 9 | 68 | 11.9 | 1676 | 4 | US-08-484-105-21 |
| 10 | 68 | 11.9 | 1676 | 4 | US-08-484-106-21 |
| 11 | 67.5 | 11.8 | 2196 | 4 | US-09-149-476-289 |
| 12 | 67.5 | 11.8 | 999 | 6 | 5514567-5 |

| Result | Sequence | Alignment | Sequence |
|--------|----------|-----------|----------|
| 13 | 67.5 | 11.8 | 4379 |
| 14 | 67.5 | 11.8 | 4379 |
| 15 | 67 | 11.7 | 2143 |
| 16 | 67 | 11.7 | 3110 |
| 17 | 66.5 | 11.6 | 261 |
| 18 | 66.5 | 11.6 | 7295 |
| 19 | 65.5 | 11.4 | 1179 |
| 20 | 65.5 | 11.4 | 1647 |
| 21 | 65.5 | 11.4 | 2560 |
| 22 | 65.5 | 11.4 | 4107 |
| 23 | 65.5 | 11.4 | 4524 |
| 24 | 64 | 11.2 | 2276 |
| 25 | 63.5 | 11.1 | 2735 |
| 26 | 63.5 | 11.1 | 2735 |
| 27 | 63.5 | 11.1 | 11309 |
| 28 | 63 | 11.0 | 5285 |
| 29 | 63 | 11.0 | 5285 |
| 30 | 62.5 | 10.9 | 10091 |
| 31 | 62 | 10.8 | 28804 |
| 32 | 62 | 10.8 | 28804 |
| 33 | 62 | 10.8 | 28804 |
| 34 | 61.5 | 10.7 | 954 |
| 35 | 61.5 | 10.7 | 954 |
| 36 | 61.5 | 10.7 | 988 |
| 37 | 61.5 | 10.7 | 988 |
| 38 | 61.5 | 10.7 | 1251 |
| 39 | 61.5 | 10.7 | 1251 |
| 40 | 61.5 | 10.7 | 1251 |
| 41 | 61.5 | 10.7 | 1345 |
| 42 | 61.5 | 10.7 | 1353 |
| 43 | 61.5 | 10.7 | 1353 |
| 44 | 61.5 | 10.7 | 1353 |
| 45 | 61.5 | 10.7 | 2191 |

ALIGNMENTS

RESULT 1
US-08-741-134-1
Sequence 1, Application US/08741134
Patent No. 5861498
GENERAL INFORMATION:
APPLICANT: Litwack, Gerald
APPLICANT: Alnemri, Emad S.
TITLE OF INVENTION: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
NUMBER OF SEQUENCES: 6
METHODS OF USING THE SAME
CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5861498-its
STREET: One Liberty Place - 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 3.11
SOFTWARE: Wordperfect for Windows 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741.134
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007.163
FILING DATE: 01-NOV-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-2090

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 160..1398
US-08-741-134-1

Alignment Scores:
Pred. No.: 0.187 Length: 2255
Score: 81.50 Matches: 33
Percent Similarity: 43.12% Conservative: 14
Best Local Similarity: 30.28% Mismatches: 34
Query Match: 14.22% Indels: 28
DB: Gaps: 5

US-09-936-737a-2 (1-103) x US-08-741-134-1 (1-2255)
QY 1 GlnGluArgLysCysTyrPheThrAlaAsnArgLysTyrThrAspPheAspLys 20
   |||||
DB 525 GAGGAACTCGAAGATGCCAATGATGCCACGCTACACAAAGGCCAGCTGACAGAA 584
QY 20 sSerPheLysSer-----SerAspLeuAspLysLysLysLysLysLys 35
   |||||
DB 585 AGCTGGCAAGACAGATGACGACGACGACGAGAAAGTATGATGATGATGATGAT 636
QY 35 sPheLysThrLysTyrLysTyrLysLysLysLysLysLysLysLysLysLys 55
   |||||
DB 637 -----GAGACCGACCTCAAG----- 654
QY 55 AsnValAlaAspLysLysLysLysLysLysLysPheValValAspLysLysPhe 75
   |||||
DB 655 -----TTCTTCGCGGTGAAGATATAGACACGATGAA-----AATGATGATCATTC 704
QY 75 rGluAsnTyrLeuThrAspCysGluGlyLysAspAlaGlyAsnAlaLacLysThrGly 95
   |||||
DB 705 AATGAAC-----ACATCAGCTGAAGGAGGACAGATGATGAAGAGATGATGAAGA 758
QY 95 pGluSerAspGluValAspGluAsp 103
   |||||
DB 759 CGAAGAGATGACAGAAATGATGAT 783

RESULT 2
US-08-044-547-4
; Sequence 4, Application US/08044547
; Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547

```

```

FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-044-547-4

Alignment Scores:
Pred. No.: 0.0232 Length: 378
Score: 80.50 Matches: 18
Percent Similarity: 50.91% Conservative: 10
Best Local Similarity: 32.73% Mismatches: 24
Query Match: 14.05% Indels: 3
DB: Gaps: 1

US-09-936-737a-2 (1-103) x US-08-044-547-4 (1-378)
QY 5 AspCysTyrPheThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 24
   |||||
DB 133 GATTCGTGCTTAACAGCGCTGCTGGAAGCGCTGACAAATCTGTGACGAAGACTGAA 192
QY 25 SerSerAspLeuAspLysLysLysLysLysLysLysPheLysThrLysLysLysLys 41
   |||||
DB 193 TTCACGCGGTGATGATGAAGCAAGAAAGATGTCGAAGAAATCGCCGTGAACATCTGCG 252
QY 42 TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsn 56
   |||||
DB 253 TACATTCGCAATCAACACAGAAACCAACGATGCTATAGAAAT 297

RESULT 3
US-08-044-547-2
; Sequence 2, Application US/08044547
; Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990

```

ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-044-547-2

Alignment Scores:
Pred. No.: 0.0285 Length: 441
Score: 80.50 Matches: 18
Percent Similarity: 50.91% Conservative: 10
Best Local Similarity: 32.73% Mismatches: 24
Query Match: 14.05% Indels: 3
DB: 1 Gaps: 1

US-09-936-737A-2 (1-103) x US-08-044-547-2 (1-441)
QY 5 AspCysTriPheThyAlaAsnArgLysTyrThrAspPheAspLysSerPheLysLys 24
DB 196 GATTGCTGCTTAACGGCGCTGCTGAAGCTGCTGCACATCTGTGCGAAGACTGAA 255
QY 25 SerSerAspLeuAspLysLysLysLysLysPheLysThrLysLysLysLysLysLys 41
DB 256 TTCACACGCGCTCATGATGAAAGATGTCGAAGATCGCCGTGGAACCATCTGC 315
QY 42 TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrPheAsn 56
DB 316 TACATCTGCAATCAACACAGAACCAACGATCTATAGAAAT 360

RESULT 4
US-08-258-287B-56
Sequence 56, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
APPLICANT: Mura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
APPLICATION NUMBER: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 45..1301
US-08-258-287B-56

Alignment Scores:
Pred. No.: 0.635 Length: 1402
Score: 75.50 Matches: 26
Percent Similarity: 54.79% Conservative: 14
Best Local Similarity: 35.62% Mismatches: 14
Query Match: 13.18% Indels: 19
DB: 3 Gaps: 6

US-09-936-737A-2 (1-103) x US-08-258-287B-56 (1-1402)
QY 44 ValPheGluAspThrValAsnLysGluCysTyrTyrAsnValAlaAspGluGluLeu 63
DB 120 GTTTTGATGATCCTGCTGCGAGAG-----AATGTTTAAATGAGATGAGTTA 167
QY 64 -----AspIleGluLysPheValAlaAsp-----GluAspPheThrGluAsn 77
DB 168 CTAATAATAGGGGAAAGTCGAGTTTCATCGAACAGGCTGAGAAATCTGTTGAGAAC 227
QY 78 TyrLeu-----ThrAspCysGluGluLys-----AspAlaGlyAsnAlaIleGlyThrGly 94
DB 228 TTCCTTAGAGAAAACAGACATGCGAGAAAATATTGCTGCGCACATTTGCCAATTTCCAC 287
QY 95 AspGlu-----SerAspGluValAsp 101
DB 288 GAACAGCTGATTTACAAATTTCTAATGATGAGATGAT 326

RESULT 5
US-08-368-704C-54
Sequence 54, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
APPLICANT: Mura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
APPLICATION NUMBER: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
APPLICATION NUMBER: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brooks, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
US-09-961-527-71

Alignment Scores:
Pred. No.: 280 Length: 32768
Score: 69.50 Matches: 25
Percent Similarity: 39.62% Conservative: 17
Best Local Similarity: 23.58% Mismatches: 35
Query Match: 12.13% Indels: 29
DB: 4 Gaps: 5

US-09-936-737A-2 (1-103) x US-08-961-527-71 (1-32768)

QY 7 TrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLysSerSer 26
Db 7201 TGGATTAGCAATTAATTAATAGCCGCTTAGAGAACTTTGAC-----TCA 7242

QY 27 AspleuasplucysLysLysThr-----CysPheLysThrGluTyrGly 41
Db 7443 GAAATTTTATTATGTCGCGAATCTGAGCAAAATTCATGTTATTAATCGGATATTAGT 7302

QY 42 TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValAlasplu 61
Db 7303 TACCACTCTT-----TTACGCTATTTCATACCTATTTCGTG 7341

QY 62 GluLeuAspGlnLysPhe-----ValValaspluAsnPhethrGlu 76
Db 7342 CAAGAGACCAAGCCCTCTCTGACCTGATCTGATCTGTAACGAAATCTGATGAC 7401

QY 77 AsnTyrIleuThrAspCysGluGly-----LysAspAlaGlyAsnAla 90
Db 7402 TTGTTGCTACAGACTTAACAGATTATCCTTTGGCTGCTGTAGAGATTTGGGGCAGA 7461

QY 91 AlaGlyThrGlyAspGlu 96
Db 7462 GCTTATTTTGCTACAGA 7479

RESULT 8
US-09-149-476-151/c
Sequence 151, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974

| | |
|---------------------------------|---|
| EARLIER FILING DATE: 1997-06-06 | EARLIER APPLICATION NUMBER: 60/056, 886 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 877 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 889 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 893 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 630 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 878 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 662 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 872 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 882 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 637 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 903 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 888 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 879 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 880 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 894 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 911 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 636 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 874 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 910 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 864 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 631 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/056, 845 |
| EARLIER FILING DATE: 1997-08-22 | EARLIER APPLICATION NUMBER: 60/047, 599 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 588 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 585 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 586 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 550 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 594 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 589 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 553 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/047, 614 |
| EARLIER FILING DATE: 1997-05-23 | EARLIER APPLICATION NUMBER: 60/043, 578 |
| EARLIER FILING DATE: 1997-04-11 | EARLIER APPLICATION NUMBER: 60/043, 576 |
| EARLIER FILING DATE: 1997-04-11 | |

| | |
|----|--|
| 1 | EARLIER APPLICATION NUMBER: 60/0447, 501 |
| 2 | EARLIER FILING DATE: 1997-05-23 |
| 3 | EARLIER APPLICATION NUMBER: 60/043, 670 |
| 4 | EARLIER FILING DATE: 1997-04-11 |
| 5 | EARLIER APPLICATION NUMBER: 60/056, 632 |
| 6 | EARLIER FILING DATE: 1997-08-22 |
| 7 | EARLIER APPLICATION NUMBER: 60/056, 664 |
| 8 | EARLIER FILING DATE: 1997-08-22 |
| 9 | EARLIER APPLICATION NUMBER: 60/056, 876 |
| 10 | EARLIER FILING DATE: 1997-08-22 |
| 11 | EARLIER APPLICATION NUMBER: 60/056, 881 |
| 12 | EARLIER FILING DATE: 1997-08-22 |
| 13 | EARLIER APPLICATION NUMBER: 60/056, 909 |
| 14 | EARLIER FILING DATE: 1997-08-22 |
| 15 | EARLIER APPLICATION NUMBER: 60/056, 875 |
| 16 | EARLIER FILING DATE: 1997-08-22 |
| 17 | EARLIER APPLICATION NUMBER: 60/056, 862 |
| 18 | EARLIER FILING DATE: 1997-08-22 |
| 19 | EARLIER APPLICATION NUMBER: 60/056, 887 |
| 20 | EARLIER FILING DATE: 1997-08-22 |
| 21 | EARLIER APPLICATION NUMBER: 60/056, 908 |
| 22 | EARLIER FILING DATE: 1997-08-22 |
| 23 | EARLIER APPLICATION NUMBER: 60/048, 964 |
| 24 | EARLIER FILING DATE: 1997-06-06 |
| 25 | EARLIER APPLICATION NUMBER: 60/057, 650 |
| 26 | EARLIER FILING DATE: 1997-09-05 |
| 27 | EARLIER APPLICATION NUMBER: 60/056, 884 |
| 28 | EARLIER FILING DATE: 1997-08-22 |
| 29 | EARLIER APPLICATION NUMBER: 60/057, 669 |
| 30 | EARLIER FILING DATE: 1997-09-05 |
| 31 | EARLIER APPLICATION NUMBER: 60/049, 610 |
| 32 | EARLIER FILING DATE: 1997-06-13 |
| 33 | EARLIER APPLICATION NUMBER: 60/061, 060 |
| 34 | EARLIER FILING DATE: 1997-10-02 |

| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 7.37 |
| Score: | 68.00 |
| Percent Similarity: | 40.00% |
| Best Local Similarity: | 25.71% |
| Query Match: | 11.87% |
| DB: | 4 |
| Length: | 1540 |
| Matches: | 27 |
| Conservative: | 15 |
| Mismatches: | 41 |
| Indels: | 22 |
| Gaps: | 5 |

US-09-936-737A-2 (1-103) x US-09-149-476-151 (1-1540)

9 PheTyrAlaAsnArgLysTyr-----ThrAspPheasp 19

Db 1111 TTTGGGCAATCGGAATACATAATGTGGCCATTGCCACAAGGAGACTGAAATCAAT 1052

QY 20 LysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys----- 3/

D6 1051 ACGAGCAAGCCAAAAGCTTCAGGATTGGA--GCCAGGATGACCAIGATGAAATGCGGA 33

| QY | Initial | Final |
|----|------------|------------|
| 38 | 1111111111 | 1111111111 |

[illegible]

878

60 valvalaenc]uasntrileuthrasn[vsG]ug[vlvsaspa]agly 88

877 CTAGCAGCATGAGCGCTTGC A A CAGCAGACTGCAGAGATTATGGCCACAGATTTCAGSA 811

89 AsnAlaIaGlyThr 93

Db 817 CTTGCTCCGGACA 803

RESULT 9

; Sequence 21, Application US/08484105


```

: GENERAL INFORMATION:
: APPLICANT: STILLMAN, Bruce
: APPLICANT: BELL, Stephen P
: APPLICANT: KOBAYASHI, Ryuji
: APPLICANT: RINE, Jasper
: APPLICANT: FOSS, Margit
: APPLICANT: MCNALLY, Francis J
: APPLICANT: LAURENSEN, Patricia
: APPLICANT: HERSKOWITZ, Ira
: APPLICANT: LI, Joachim J
: APPLICANT: GAVIN, Kimberly
: TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,105
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard Aron
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEEX: 910 277299
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..1302
: US-08-484-105-21
:
: Alignment Scores:
: Pred. No.: 8.26 Length: 1676
: Score: 68.00 Matches: 29
: Percent Similarity: 45.54% Conservative: 22
: Best Local Similarity: 25.89% Mismatches: 24
: Query Match: 11.87% Indels: 37
: Gaps: 7
:
: US-09-936-737a-2 (1-103) x US-08-484-105-21 (1-1676)
:
: QY 20 LysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys----- 37
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 24 AAAAATTTTGAACGACGACTGTCACGCGCCAGTCCGCGTTCGTGAAAAAATCGAC 83
:
: QY 38 -----ThnGluTyrCysTyrIleValbheGluAspThrValAsn 50
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 84 TCCGAAAAAAGAGATCCAGACAGAA-----AAAGACGAATGGA 125
:
: QY 51 LysGluCysTyr-TyrAsnValValAspGly-----GluGluLeuAspGln----- 65
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 126 AGAAGATGCTTGTAGAAATTTGCAATCAATTTAGAAAGAGATTGGAACAATCGGCTT 185
:
: QY 66 -----GluLysPheValValAspGluAsnPhetThnGluAspTyrIleuThrAspCysG 83
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 186 CGAGATGAACCTGTATCAATGCGTCATCAGCATCGAAAAATTACTT-----ATGCA 239
:
: QY 83 uGlyLysAspAla-----GlyAsnAlaAla 92
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 240 AGGAAAAATCGCGCTCAGAACGAAATGAATATGCGAAATCCGCTCGGAAGACGTCTGG 299
:
: QY 92 yThrGlyAspGluSerAspGluValAspGluAsp 103
: ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 300 AAATGGAAAT--ACTGAAGAAATTTGAGGAAGAC 330
:
: RESULT 10
: US-08-484-106-21
: Sequence 21, Application US/08484106
: Patent No. 5614518
:
: GENERAL INFORMATION:
: APPLICANT: STILLMAN, Bruce
: APPLICANT: BELL, Stephen P
: APPLICANT: KOBAYASHI, Ryuji
: APPLICANT: RINE, Jasper
: APPLICANT: FOSS, Margit
: APPLICANT: MCNALLY, Francis J
: APPLICANT: LAURENSEN, Patricia
: APPLICANT: HERSKOWITZ, Ira
: APPLICANT: LI, Joachim J
: APPLICANT: GAVIN, Kimberly
: TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,106
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard Aron
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEEX: 910 277299
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..1302
: US-08-484-106-21
:
: Alignment Scores:
: Pred. No.: 8.26 Length: 1676
: Score: 68.00 Matches: 29
: Percent Similarity: 45.54% Conservative: 22
: Best Local Similarity: 25.89% Mismatches: 24
: Query Match: 11.87% Indels: 37
: Gaps: 7
:
: US-09-936-737a-2 (1-103) x US-08-484-106-21 (1-1676)

```

```
QY 20 LysSerPheLysSerSerAspLeuAspGluCysLysLysThrCysPheLys----- 37
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 24 AAAAATTTGAACGAGCACTGTCACGCGCCGCGCTTCTTGA AAAAATGAC 83
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 38 -----ThGluTyrCysTyrIleValPheGluAspThrValAsn 50
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 84 TCAGAAAAAGAGATCCAGACAGAA-----AAAGCAATGAGAA 125
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 51 LysGluCysTyr-TyrAsnValValAspGly-----GluGluLeuAspGln----- 65
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 126 AGAGAACTCTTGAATTTGCAATCAATTTAGAGAGAACTTGGAAACACTGGGCTT 185
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 66 -----GluLysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCysG1 83
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 186 CGAGATCAAACTGATTCATGCGTCAATCGCAATCGAAATATCTT-----ATGCA 239
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 83 uGlyLysAspAla-----GlyAsnAlaAlaG1 92
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 AGGAAATCGCGCTCAGAACGAAATGAATATCGAAATCCCGCTCGTGAAGACGCTG 299
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 92 yThrGlyAspGluSerAspGluValAspGluAsp 103
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 AATGCAAT--ACTGAAGAATTTAGGAGAC 330
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 11
US-09-149-476-289/C
Sequence 289 Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149, 476
EARLIER FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCI/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
```

[REDACTED]

1

```

; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Alignment Scores:
Pred. No.: 11.9
Score: 68.00
Percent Similarity: 40.008
Best local Similarity: 25.778
Query Match: 11.878
Db: 4
Gaps: 5

US-09-936-737A-2 (1-103) x US-09-149-476-289 (1-2196)

QY 9 PheTYrAlaAsnArgLysr-----ThraSpleasp
|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 827 TTTTGGCAAAATCGGAATACATAATGTGGCCATTGCCAAGAGAGACTGAATCAT
QY 20 LysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCys-----
||| ||| |||:||||| ||| ||| |||
Db 767 ACAGACAGCCAAAGCTTCAGATTGGA---GCCAGATGACATGATGAATGCAGA
QY 38 -----ThGluTYrCys-----TYrIleValPheGluAspThr
|||:||||| |||:||||| |||:|||||
Db 710 AGATCAAGACAGATAGTCATGAGAGCTCTGTACACCATTTATATAGCGCTTCACATTCA
QY 49 ValAsnLysGluCysTYrTYrAsnValValAspGlyGluGluIleuAspGluLysPhe
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 ATTCAATTCTTCTGC---AGCAACTGTGTACACAGTTAATCAAGACCAAGCCGATT
QY 69 ValValAspGluAsnPheThrGluAsnTYrIleuThraSpleCysGluGlyLysAspAlaGly
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 CTACGACGAATGAGCGCTGTCAACACAGACACTGACAGAGATTATGGGCACAGATTACAGA
QY 89 AsnAlaIaGlyThr 93
|||:|||||
Db 533 CTGTCTCCGGAC 519

RESULT 12
5514567-5
Patent No. 5514567
APPLICANT: SUGANO, HARUO;MURAMATSU, MASAMI;TANIGUCHI,
TADATOSUGU
TITLE OF INVENTION: DNA AND RECOMBINANT PLASMID
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/400,179
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 389,922
FILING DATE: 18-JUN-1982
APPLICATION NUMBER: 201,359
FILING DATE: 27-OCT-1980
SEO ID NO:5
LENGTH: 999
5514567-5

Alignment Scores:
Pred. No.: 4.82
Score: 67.50
Percent Similarity: 50.708
Best local Similarity: 32.398
Query Match: 11.788
Db: 6
Gaps: 2

US-09-936-737A-2 (1-103) x 5514567-5 (1-999)
QY 11 AlaAsnArgLysrThraSpleaspLysSerPheLysSerPheLysSerSerAspLeuAspGlu 30
|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Oy      16  ThrApplPheAspLysSerSphelysYsSerSerAsp-----LeuAspGluCysIlysLys 33
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1119  ACGAGTTCGTACGACGAGTTTAAAGCAAAAGACGAAGATCGAAGAAATTAATTAAAGTAT 1060
Oy      34  ThCysPheLysrThGluTyrCysTyrTLeValPheGluAspThrValAsnLysGluCys 53
          :::  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1059  TCAAAATTTCAAAACCCAAAATAACTATTGTC---AAAGATTTTTCATCATGTTGTGTCT 1007
Oy      54  TTYrAsnValValAspGlyGluGluLeuAspGln 65
          |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1002  TCTATTAGTGTGTAGATGATCAGACAAAAAATCAG 967

RESULT 14
US-09-149-976-17/c
; Sequence 17, Application US/09149976
; Patent No. 6127123
GENERAL INFORMATION:
APPLICANT: Yanofsky, Martin F.
TITLE OF INVENTION: Cauliflower Floral Meristem Identity
TITLE OF INVENTION: Genes and Methods of Using Same
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/149,976
FILING DATE: 09-SEP-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,214
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 3291
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 4379 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: unsure
LOCATION: 2095..2098
OTHER INFORMATION: /note= "N = one or more
OTHER INFORMATION: nucleotides."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..4379
OTHER INFORMATION: /note= "sequence = Arabidopsis
OTHER INFORMATION: thaliana API gene"
US-09-149-976-17

Alignment Scores:
Pred. No.: 35 Length: 4379
Score: 67.50 Matches: 19
Percent Similarity: 53.85% Conservative: 9
Best Local Similarity: 36.54% Mismatches: 21
Query Match: 11.78% Indels: 3
DB: 3 Gaps: 2

```

US-09-936-737A-2 (1-103) x US-09-149-976-17 (1-4379)

QY 16 ThrAspPheAspLysSerPheLysSerSerAsp-----LeuAspGluCysLysLys 33
DB 1119 ACGGATTTTACGACGAGATTAAAGAGAGAGAAAGATCGAGAGATTAATTGTAAGTAT 1060
QY 34 ThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThrValAsnLysGluCys 53
DB 1059 TCATAATTTCACAAACCAAAATAGTATTTTTCG---AAAGATTTTTCATGATGTTGTGTGT 1003
QY 54 TyrTyrAsnValValAspGluGluLeuAspGln 65
DB 1002 TCTTATTAAGTTGATGATGATGAGACAAATATCAG 967

RESULT 15

US-09-069-023-21
; Sequence 21, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; FILE REFERENCE: UM-03333
; CURRENT APPLICATION NUMBER: US/09/069, 023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2143
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-069-023-21

Alignment Scores:

| | | | |
|------------------------|--------|---------------|------|
| Pred. No.: | 15.7 | Length: | 2143 |
| Score: | 67.00 | Matches: | 32 |
| Percent Similarity: | 43.36% | Conservative: | 17 |
| Best Local Similarity: | 28.32% | Mismatches: | 28 |
| Query Match: | 11.69% | Indels: | 36 |
| DB: | 4 | Gaps: | 9 |

US-09-936-737A-2 (1-103) x US-09-069-023-21 (1-2143)

QY 14 LysTyrThrAspPheAspLysSerPheLys-----SerSerAspLeuAspGlu 30
DB 278 AAAGTCAAGGACGTAACACCAATTAAGAGAGCTGTCTGCGCTCCAGCTCGAGAGAG 337
QY 31 CysLys---LysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThrVal 49
DB 338 ATCCGCGAGGAGTGGCGGAGAGTTCGAGAGTGGACCACTG-----CCCAACATC 391
QY 50 AsnLysGluCysTyrTyrAsnValValAspGluGluLeuAspGlnGluLysPhe--- 68
DB 392 CACCTGACCTCC-----GACGCGACGAGATGATGATGAGGAGTACTTT 436
QY 69 ---ValValAspGluAsnPheThrGlu----- 76
DB 437 CGACACCTCGACGAGAAAT---ACGGAACGTGGTCCGCTTTCCCGAGAAACATTTGATC 493
QY 77 -----AsnTyr-----LeuThrAspCysGluGlyLysAspAlaGly----- 88
DB 494 GATCCACCTACTACTGAGTAAACCACTCCACATGCAACGAGGAGGAACTGGAAC 553
QY 89 -----AsnAlaAlaGlyThrGlyAspGluSerAsp 98
DB 554 GGAGAGCTGAGCGAGGAGCGAGGCGCACACCGAGT 592

Search completed: February 25, 2003, 03:23:14
Job time : 53 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:34:15 ; Search time 62 Seconds

(without alignments)
933.046 Million cell updates/sec

Title: US-09-936-737A-2

Sequence: 1 EEREDCMTFYANKRYTDFDK.....GMDAGNAGTGDSEDEVED 103

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 442118 segs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+pn.model -DEV=xlh
-Q/cg2_1/USPTO.spool/US0936737/runat_14022003_100733_2323/app_query.fasta_1.263
-DB=Published.Applications.NA -QEMT=fastap -SUFFIX=trpb -MIMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MIMLEN=0
-MIMLEN=200000000 -USER=US0936737 -RCGN_1_1_33_etunat_14022003_100733_2323
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database: Published.Applications.NA:

1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/1/pubpna/PCCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/1/pubpna/PCCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 71.5 | 12.5 | 1092 | 10 | US-09-815-242-9446 |
| 2 | 70.5 | 12.3 | 543 | 9 | US-09-796-692-9477 |
| 3 | 70 | 12.2 | 540 | 9 | US-10-040-739-838 |
| 4 | 70 | 12.2 | 3504 | 10 | US-09-815-242-7425 |

| | | | | | | | |
|---|----|------|------|--------|----|---------------------|--------------------|
| C | 5 | 70 | 12.2 | 197997 | 10 | US-09-822-246-3 | Sequence 3, Appl1 |
| C | 6 | 69.5 | 12.1 | 2023 | 10 | US-09-765-272-199 | Sequence 199, App |
| C | 7 | 68 | 11.9 | 1861 | 10 | US-09-729-674-129 | Sequence 129, App |
| C | 8 | 68 | 11.9 | 1988 | 10 | US-09-864-761-4740 | Sequence 4740, Ap |
| C | 9 | 68 | 11.9 | 5153 | 10 | US-09-727-384-3 | Sequence 3, Appl1 |
| C | 10 | 67.5 | 11.8 | 914 | 9 | US-09-798-889-29 | Sequence 29, Appl |
| C | 11 | 67.5 | 11.8 | 2664 | 10 | US-09-834-975-897 | Sequence 897, App |
| C | 12 | 67 | 11.7 | 763 | 10 | US-09-910-943-179 | Sequence 179, App |
| C | 13 | 67 | 11.7 | 1430 | 9 | US-09-759-1308-61 | Sequence 61, Appl |
| C | 14 | 67 | 11.7 | 2330 | 9 | US-08-759-1308-51 | Sequence 51, Appl |
| C | 15 | 67 | 11.7 | 3110 | 10 | US-09-939-980-237 | Sequence 237, App |
| C | 16 | 66.5 | 11.6 | 3984 | 9 | US-10-149-110-2 | Sequence 2, Appl1 |
| C | 17 | 66.5 | 11.6 | 7450 | 9 | US-10-149-110-10 | Sequence 10, Appl |
| C | 18 | 66.5 | 11.6 | 8370 | 10 | US-09-801-574-56 | Sequence 56, Appl |
| C | 19 | 66.5 | 11.6 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appl1 |
| C | 20 | 66 | 11.5 | 634 | 10 | US-09-770-149-663 | Sequence 663, App |
| C | 21 | 66 | 11.5 | 5988 | 10 | US-09-864-864-281 | Sequence 281, App |
| C | 22 | 66 | 11.5 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appl1 |
| C | 23 | 65.5 | 11.4 | 2155 | 10 | US-09-960-428-13 | Sequence 13, Appl |
| C | 24 | 65 | 11.3 | 436 | 10 | US-09-867-701-1512 | Sequence 1512, Ap |
| C | 25 | 65 | 11.3 | 3611 | 9 | US-09-974-298-113 | Sequence 113, App |
| C | 26 | 64.5 | 11.3 | 288 | 10 | US-09-974-300-3913 | Sequence 3913, Ap |
| C | 27 | 64.5 | 11.3 | 469 | 10 | US-09-864-761-6338 | Sequence 6338, App |
| C | 28 | 64.5 | 11.3 | 562 | 10 | US-09-954-456-216 | Sequence 216, App |
| C | 29 | 64.5 | 11.3 | 562 | 10 | US-09-954-456-822 | Sequence 822, App |
| C | 30 | 64.5 | 11.3 | 562 | 10 | US-09-954-456-1225 | Sequence 1225, Ap |
| C | 31 | 64.5 | 11.3 | 630 | 10 | US-09-938-842A-2486 | Sequence 2486, App |
| C | 32 | 64.5 | 11.3 | 1152 | 9 | US-10-042-417-51 | Sequence 51, Appl |
| C | 33 | 64 | 11.2 | 1777 | 12 | US-09-822-846-421 | Sequence 421, Appl |
| C | 34 | 64 | 11.2 | 2807 | 10 | US-09-940-836A-1 | Sequence 3, Appl1 |
| C | 35 | 64 | 11.2 | 2843 | 9 | US-09-900-038A-3 | Sequence 4933, Ap |
| C | 36 | 64 | 11.2 | 6865 | 10 | US-09-878-574-4933 | Sequence 4933, Ap |
| C | 37 | 63.5 | 11.1 | 420 | 10 | US-09-878-574-4933 | Sequence 1717, Ap |
| C | 38 | 63.5 | 11.1 | 1629 | 9 | US-09-738-626-1717 | Sequence 1, Appl |
| C | 39 | 63.5 | 11.1 | 1839 | 10 | US-09-931-768A-1 | Sequence 1380, Ap |
| C | 40 | 63.5 | 11.1 | 2080 | 9 | US-09-938-842A-1380 | Sequence 7, Appl |
| C | 41 | 63 | 11.0 | 1839 | 10 | US-09-930-871-7 | Sequence 17, Appl |
| C | 42 | 63 | 11.0 | 4166 | 10 | US-09-930-871-17 | Sequence 19, Appl |
| C | 43 | 63 | 11.0 | 4179 | 10 | US-09-930-871-17 | |
| C | 44 | 63 | 11.0 | 4179 | 10 | US-09-930-871-17 | |
| C | 45 | 63 | 11.0 | 4197 | 10 | US-09-930-871-19 | |

ALIGNMENTS

RESULT 1
US-09-815-242-9446/c
; Sequence 9446, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

```

US-09-936-737A-2 (1-103) x US-09-815-242-9446 (1-1092)

? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 9446
? LENGTH: 1092
? TYPE: DNA
? ORGANISM: Streptococcus pneumoniae
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)...(1092)
US-09-815-242-9446

Alignment Scores:
Pred. NO.: 1.43
Score: 71.50
Percent Similarity: 42.05%
Best Local Similarity: 27.27%
Query Match: 12.48%
DB: 10
Gaps: 4

Length: 1092
Matches: 24
Conservative: 13
Mismatch: 32
Indels: 19
Gaps: 4

```

```

? PRIOR FILING DATE: 2000-05-22
? PRIOR APPLICATION NUMBER: 60/218,950
? PRIOR FILING DATE: 2000-07-14
? PRIOR APPLICATION NUMBER: 60/222,903
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: 60/223,416
? PRIOR FILING DATE: 2000-08-04
? PRIOR APPLICATION NUMBER: 60/223,378
? PRIOR FILING DATE: 2000-08-07
? NUMBER OF SEQ ID NOS: 9597
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 9477
? LENGTH: 543
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (436)
? OTHER INFORMATION: n=A,T,C or G
? NAME/KEY: unsure
? LOCATION: (485)
? OTHER INFORMATION: n=A,T,C or G
? US-09-796-692-9477

```

```

QY      6  CysrpttrtrphtyAlaAsnArgLysThrThrAspPheAspLysSerPheLysSer 25
      |||||
Db      1065 TGTGTGAGC-----AACCATCCACGTTTCATGAAGATTTC----- 1030

QY      26  SerAspLeuaspGluCysLysLysThrCysPheLys----- 37
      |||||
Db      1029 AGCATCTGACATGATGTTCCAAAGCAGCACCTTTGAGAGGTTATCTGAACAACACCAT 970

QY      38  -----ThrgLutryCysTyrIleValPheGluAspThrValAsnLysGluCysTyr--- 54
      |||||
Db      969 GTGAATTCCTTTTCTGTCATCCAAAGTCTTACGGATACGACCAACAAAGGTATCACGTGA 910

QY      55  TyrAsnValValAspGluGluLeuAspGlnGluLysPheValValAspGluAspPhe 74
      |||||
Db      909 ACCAACTGCATGTATGCGTTGAGATAGATTGTGATGAACATCATCATCTTCAAGACAGC 850

QY      75  ThrGluAsnTyrLeuThrAspCys 82
      |||||
Db      849 ACCTGGAGAGCATGCGATAGCTGC 826

RESULT 2
US-09-796-692-9477
: Sequence 9477, Application US/09796692
: Publication No. US20020198362A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
: FILE REFERENCE: 2077_001200
: CURRENT APPLICATION NUMBER: US/09/796,692
: CURRENT FILING DATE: 2001-03-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: 60/200,201

```

| | | | |
|------------------------|--------|---------------|-----|
| Pred. No.: | 0.746 | Length: | 543 |
| Score: | 70.50 | Matches: | 26 |
| Percent Similarity: | 38.66% | Conservative: | 20 |
| Best Local Similarity: | 21.85% | Mismatches: | 50 |
| Query Match: | 12.30% | Indels: | 23 |
| DB: | 9 | Gaps: | 6 |

US-09-936-737A-2 (1-103) x US-09-796-692-9477 (1-543)

| | | | | |
|----|-----|---|--|-----|
| QY | 1 | GLUGLUARGLU----- | -ASPCYSTRPPhrPheTYrAlaasnArgLYsTYr | 15 |
| | | ::: | | ::: |
| DB | 97 | GAAGAACACGAAAAATTA | CTGAGAAAAACGTGACGTATATGTGGAAATCTTCTTTT | 15 |
| QY | 16 | ThAsPheAspLYsSer----- | -PhelysLYsSerAsp----- | 27 |
| | | ::::: | | |
| DB | 157 | TACCACTGAAGAACAAATCT | ATGACCTCTTCAGCAAAAGTGTCACATTAAGAAATATC | 21 |
| QY | 28 | -----LeaAspGLUCysLYsLYsThrCYsPheLYsThrGLUArgCYsTYrIleVal | 44 | |
| | | ::: | | ::: |
| DB | 217 | ATTATGGCTGTGGATRAAATG | AGAAAAACAGCA-----TGCGATTCTGTTTGTCGAA | 27 |
| QY | 45 | PhcGLUAsPThrValAsnLYsGLUCysTYrTYrAsnValAlaSpelYGLUleuAsp | 64 | |
| | | ::: | | |
| DB | 271 | TATTACTCAGCGCGAGATCGG | AAACCCCATGCGGTACATTAATAGCAACGGCTCGGAT | 33 |
| QY | 65 | GLNGLUAsPheValAlaAspGLUAsnPhcThrGLUAsnTYrIleuThrasPCysGLUly | 84 | |
| | | ::::: | | |
| DB | 331 | GACCGA-----ATCATTTCCGACAGACAGCTGGACGACGCTTTAAG----- | -GAGGCG | 37 |
| QY | 85 | LYsAspAlaGLYAsnAlaGLYThrcGLYAspLYsSerAspGLUValAspGLUAsp | 103 | |
| | | ::: | | |
| DB | 376 | AGGCAATACGCCCGTGGCGCAT | GTGGGGCCAGGTTGGGATGATATCGCAGGAC | 432 |

RESULT 3
US-10-040-739-838
Sequence 838, Application US/10040739
Patent No. US20020173635A1

GENERAL INFORMATION:
APPLICANT: JACOBS, Kenneth
McCoy, John
Lavallee, Edward
Racie, Lisa
Merberg, David
treacy, Maurice
Spaulding, Vikki

TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 519
CORRESPONDENCE ADDRESS:


```

ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998

ATTORNEY/AGENT INFORMATION:
NAME: BROWN, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 838:
SEQUENCE CHARACTERISTICS:
LENGTH: 540 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 838:
US-10-040-739-838

Alignment Scores:
Pred. No.: 0.866 Length: 540
Score: 70.00 Matches: 31
Percent Similarity: 43.75% Conservative: 18
Best Local Similarity: 27.68% Mismatches: 50
Query Match: 12.22% Indels: 13
DB: Gaps: 5

US-09-936-737a-2 (1-103) x US-10-040-739-838 (1-540)
QY 2 GUAAGGUASPCYSTRPThrPheTyr-----AlaAsnArgLysTyrThrAsp-PH 18
DB 204 GAAAAAACAATGTTCAACATCTACAGACTCTTCCTCGAAGAAAAAACAAGACATT 263
QY 18 eAspLysSerPheLysLys-----SerSerAspLeuAspLysLysLysLysThrCysPh 36
DB 264 AGACTCAGGACCTCTGAAATTTGAAATCTCCAGAAATCGAGTGTCTTAAGACA----- 318
QY 36 eLysThrGluTyrCysTyrIleVal---PheGluAspThrValAsnLysGluCysTyrTyr 55
DB 319 -AGAGAGAAATGCAATCGATGTTCAACTCATATATAGCAAGACAGCTGTTTGA 377
QY 55 rAsnValValAspGluGluGluLeuAspGlnGluLysPheValValAspGluAsnPherh 75
DB 378 AAACCTCAACACATGATGTAATGTGTGACAAAGATCTCTGAGGAGAAATTCAGAAAGCGT 437
QY 75 rGluAsnTyrLeuThrAspCysGluGluLysAsp-----AlaGluAsnAlaAl 91
DB 438 TAGTGAATAGCAAGTGTATAGGAATCTGAAATGAATGAATTAATCAAGTGTGTGAGAGCTTC 497
QY 91 aGlyThrGluYAspGluSerAspGluValAspGlu 102
DB 498 AGGTGATGACGATGGAAGTGAATGATGAAGAG 531

```

```

RESULT 4
US-09-815-242-7425
; Sequence 7425 Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

```

```

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7425
LENGTH: 3504
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(3504)
US-09-815-242-7425

Alignment Scores:
Pred. No.: 11.5 Length: 3504
Score: 70.00 Matches: 24
Percent Similarity: 46.15% Conservative: 12
Best Local Similarity: 30.77% Mismatches: 24
Query Match: 12.22% Indels: 18
DB: Gaps: 4

US-09-936-737a-2 (1-103) x US-09-815-242-7425 (1-3504)
QY 19 AspLysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLysThr 38
DB 718 GATTAACAGTACGAAATCTGAAATCTCCAGAAATCGGCAAGAAATCTTTTGAAGCGT 777
QY 39 GluTyrCysTyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValVal 58
DB 778 -----ATCTTGTCGCAAGATCTAGCGAAC-----GCCGTGATTAACGTCATG 819
QY 59 AspGluGluGluLeuAspGlnGluLysPheValValAspGluAsnPherhCysLysThr 78
DB 820 CCCACGCAAA-----TTAGGGACAGCAAGATTAATTTGGGAAATTTTC 858
QY 79 LeuThrAspCysGluGluLysAspAlaGluValAsnAlaGlyThrGluYAspGlu 96
DB 859 -----GCTAATAAAACGGGCAACATCGCAAGACCTTGAATGAG 897

```

```

RESULT 5
US-09-822-246-3/c
; Sequence 3 Application US/09822246
; Patent No. US20020142383A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; AND USES THEREOF

```

TOPOLOGY: linear

Score: 68.00


```

Db 1760 TGCCTTCATGCCCTTAATAGATGATTCGATGATGAATTCAGTAATTCATGTAACCTTC 1819
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NM1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 1402
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-759-130B-61
;
Alignment Scores:
;
Pred. No.: 8.32
Score: 67.00
Percent Similarity: 42.62%
Best Local Similarity: 34.43%
Query Match: 11.69%
DB: 9
Gaps: 3
US-09-936-737a-2 (1-103) x US-09-759-130B-61 (1-1402)
;
Db 33 LysThrcysPheIysThrIuTyrcysTyrlleValPheGluaspThrValasnLys--- 51
;
Db 818 AAACATGATTTCTTAACA---TTTATCCCTGCTACATATGCCCAATATGTAATTAATAACA 762
;
Db 52 -----GluCysTyrrTyrrasnValValaspGluIuLeuasp 64
;
Db 761 GCTCAAGCTTTAAGAGGCCAGGACCTATGTAATGTTG----- 720
;
Db 65 GlngluIuPheValaspGluaspGluasnPheThrIuasnTyrlleuThraspcysGluIu 84
;
Db 719 ---GTTAAATAATGTTTAAATATAAAGGCCCTCAATATTATTATGGCCCTCAAGGC 663
;
Db 85 Lys 85
;
Db 662 AAA 660
;
RESULT 14
; Sequence 51, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Fraser, Christopher C
; APPLICANT: McCarthy, Sean A
; APPLICANT: Kirsch, Susan J
; APPLICANT: Mackey, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Barnes, Thomas S

```

```

; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NM1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 1402
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-759-130B-61
;
Alignment Scores:
;
Pred. No.: 8.32
Score: 67.00
Percent Similarity: 42.62%
Best Local Similarity: 34.43%
Query Match: 11.69%
DB: 9
Gaps: 3
US-09-936-737a-2 (1-103) x US-09-759-130B-61 (1-1402)
;
Db 33 LysThrcysPheIysThrIuTyrcysTyrlleValPheGluaspThrValasnLys--- 51
;
Db 818 AAACATGATTTCTTAACA---TTTATCCCTGCTACATATGCCCAATATGTAATTAATAACA 762
;
Db 52 -----GluCysTyrrTyrrasnValValaspGluIuLeuasp 64
;
Db 761 GCTCAAGCTTTAAGAGGCCAGGACCTATGTAATGTTG----- 720
;
Db 65 GlngluIuPheValaspGluaspGluasnPheThrIuasnTyrlleuThraspcysGluIu 84
;
Db 719 ---GTTAAATAATGTTTAAATATAAAGGCCCTCAATATTATTATGGCCCTCAAGGC 663
;
Db 85 Lys 85
;
Db 662 AAA 660
;
RESULT 14
; Sequence 51, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Fraser, Christopher C
; APPLICANT: McCarthy, Sean A
; APPLICANT: Kirsch, Susan J
; APPLICANT: Mackey, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Barnes, Thomas S

```

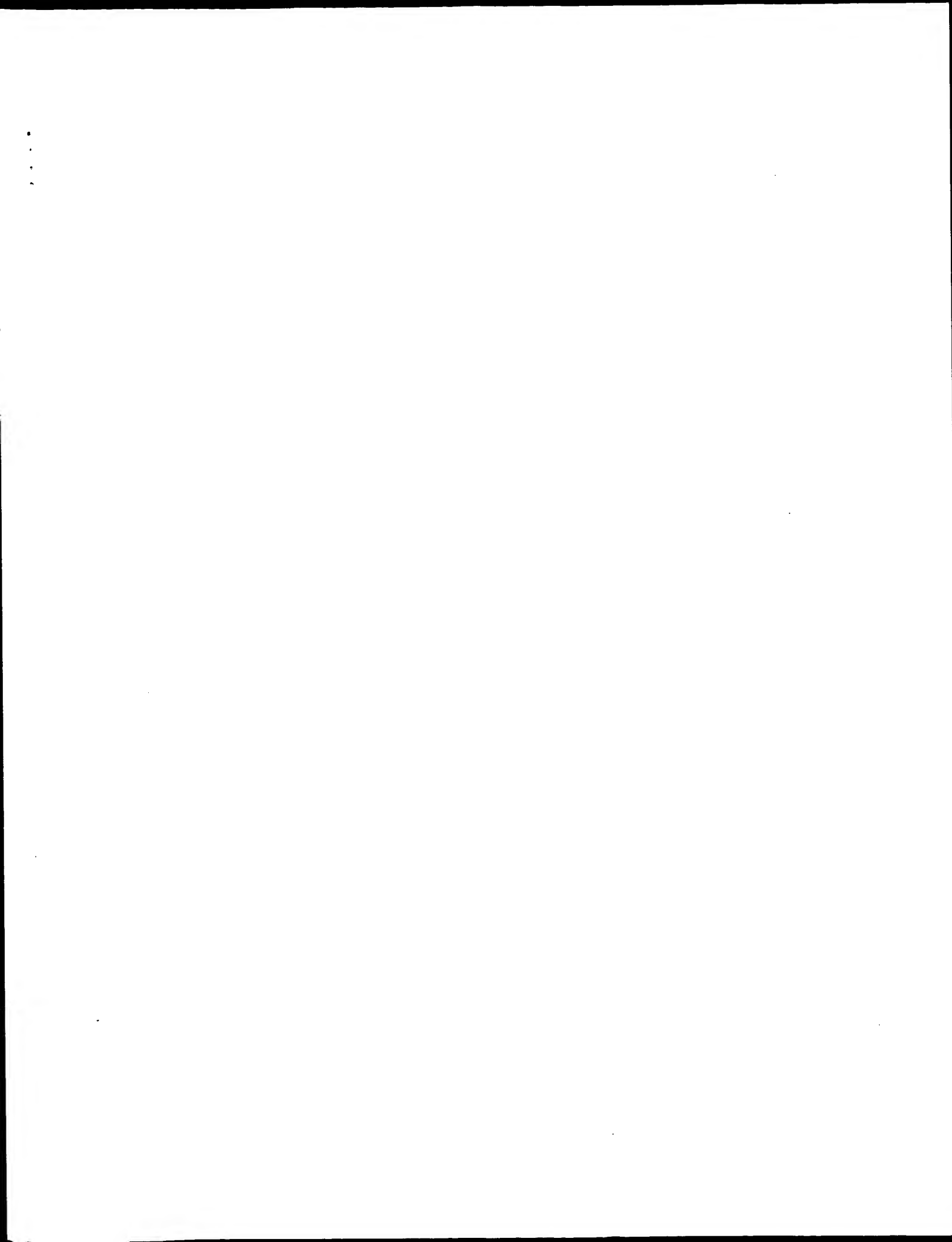
```

; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NM1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 1402
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-759-130B-61
;
Alignment Scores:
;
Pred. No.: 8.32
Score: 67.00
Percent Similarity: 42.62%
Best Local Similarity: 34.43%
Query Match: 11.69%
DB: 9
Gaps: 3
US-09-936-737a-2 (1-103) x US-09-759-130B-61 (1-1402)
;
Db 33 LysThrcysPheIysThrIuTyrcysTyrlleValPheGluaspThrValasnLys--- 51
;
Db 818 AAACATGATTTCTTAACA---TTTATCCCTGCTACATATGCCCAATATGTAATTAATAACA 762
;
Db 52 -----GluCysTyrrTyrrasnValValaspGluIuLeuasp 64
;
Db 761 GCTCAAGCTTTAAGAGGCCAGGACCTATGTAATGTTG----- 720
;
Db 65 GlngluIuPheValaspGluaspGluasnPheThrIuasnTyrlleuThraspcysGluIu 84
;
Db 719 ---GTTAAATAATGTTTAAATATAAAGGCCCTCAATATTATTATGGCCCTCAAGGC 663
;
Db 85 Lys 85
;
Db 662 AAA 660
;
RESULT 14
; Sequence 51, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Fraser, Christopher C
; APPLICANT: McCarthy, Sean A
; APPLICANT: Kirsch, Susan J
; APPLICANT: Mackey, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Barnes, Thomas S

```


Db 1357 AGTTAGATGTAGATGATACAAATGTGCTATCAAAATTCAAAGAGAGTTGAACCTAGA 1416
Qy 67 LysPheValValAspGluAsnPheThrGlu-----AsnTyrLeuThrAspCysGlu 83
|||||
Db 1417 AAGTTTGT-----TTCACCTAACGGCGAAATTAATATACAGCAATTTGNA 1464
Qy 84 GlyLysAspAlaGlyAsnAlaAlaGlyThrGlyAspGluSerAspGluValAspGluAsp 103
:::|||||
Db 1465 AAAGAAGACGTACCGAATGCAATCAAGCCAAAGAGTGAATCAGATTGTGAGTGATGAGNA 1524
:::|||||

Search completed: February 25, 2003, 03:57:32
Job time : 88 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:29:55 ; Search time 14 Seconds

(without alignments)
707.275 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EEREDCWTFFANRRKTYDFDK.....GKDAGNAAGTGDSEDEVDED 103

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR:73:*
2: PIR:1:*
3: PIR:3:*
4: PIR:4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query length | DB ID | Description |
|------------|-------|-------|--------------|-------|----------------------------|
| 1 | 80.5 | 14.0 | 147 | 2 | A42435 leech antiplatelet |
| 2 | 80.5 | 14.0 | 412 | 2 | A55320 immunophilin FKBP4 |
| 3 | 78.5 | 13.7 | 550 | 2 | T26562 hypothetical prote |
| 4 | 78.5 | 13.7 | 907 | 2 | T27317 hypothetical prote |
| 5 | 78 | 13.6 | 682 | 2 | B84021 two-component sens |
| 6 | 78 | 13.6 | 797 | 2 | A36811 hypothetical prote |
| 7 | 75.5 | 13.2 | 264 | 2 | T22054 hypothetical prote |
| 8 | 75.5 | 13.2 | 491 | 2 | T16354 hypothetical prote |
| 9 | 75 | 13.1 | 360 | 2 | T34510 hypothetical prote |
| 10 | 75 | 13.1 | 3844 | 2 | T18402 asparagine/asparta |
| 11 | 74.5 | 13.0 | 1277 | 2 | T32731 PAR interacting pr |
| 12 | 74 | 12.9 | 296 | 2 | D70465 hypothetical prote |
| 13 | 74 | 12.9 | 1278 | 2 | B70236 hypothetical prote |
| 14 | 74 | 12.9 | 4466 | 1 | S17231 dynein beta heavy |
| 15 | 73 | 12.7 | 643 | 2 | T21428 hypothetical prote |
| 16 | 73 | 12.7 | 1405 | 1 | D72PA DNA-directed DNA p |
| 17 | 73 | 12.7 | 2543 | 2 | T31687 surface antigen - p |
| 18 | 72.5 | 12.7 | 564 | 2 | T25945 alpha-51D-immobill |
| 19 | 72.5 | 12.7 | 2533 | 2 | T28675 probable Athila re |
| 20 | 72.5 | 12.7 | 2533 | 2 | T28675 Ser/Thr protein ki |
| 21 | 72 | 12.6 | 1503 | 2 | B84493 para-aminobenzoate |
| 22 | 71.5 | 12.5 | 320 | 2 | A71611 hypothetical prote |
| 23 | 71.5 | 12.5 | 594 | 2 | F81359 alpha-51D-immobill |
| 24 | 71 | 12.4 | 782 | 2 | S50719 Ser/Thr protein ki |
| 25 | 70.5 | 12.3 | 146 | 1 | T37222 phospholipase A2 (|
| 26 | 70.5 | 12.3 | 156 | 2 | T37222 dimeric cap bindin |
| 27 | 70.5 | 12.3 | 156 | 2 | S60109 nuclear cap bindin |
| 28 | 70.5 | 12.3 | 256 | 2 | G96774 hypothetical prote |
| 29 | 70.5 | 12.3 | 748 | 2 | C64409 hypothetical prote |

| | | | | | |
|----|------|------|------|---|---------------------------|
| 30 | 70.5 | 12.3 | 786 | 2 | C86406 88.6K hypothetical |
| 31 | 70 | 12.2 | 328 | 2 | T24432 hypothetical prote |
| 32 | 70 | 12.2 | 1058 | 2 | S50295 hypothetical prote |
| 33 | 70 | 12.2 | 1167 | 2 | F71909 hypothetical prote |
| 34 | 70 | 12.2 | 1255 | 2 | B97104 hypothetical prote |
| 35 | 69.5 | 12.1 | 146 | 1 | PSR6A phospholipase A2 (|
| 36 | 69.5 | 12.1 | 452 | 2 | T37899 hypothetical prote |
| 37 | 69.5 | 12.1 | 628 | 2 | S61160 hypothetical prote |
| 38 | 69.5 | 12.1 | 696 | 2 | D95206 hypothetical prote |
| 39 | 69.5 | 12.1 | 738 | 2 | A13625 ribonucleoside-dip |
| 40 | 69 | 12.0 | 164 | 2 | T32824 hypothetical prote |
| 41 | 69 | 12.0 | 374 | 2 | T33328 hypothetical prote |
| 42 | 69 | 12.0 | 1127 | 2 | T28317 ORF MSY156 hypothe |
| 43 | 69 | 12.0 | 1650 | 2 | T18444 hypothetical prote |
| 44 | 68.5 | 12.0 | 385 | 2 | T26487 hypothetical prote |
| 45 | 68.5 | 12.0 | 597 | 2 | D70100 phosphoglucomutase |

ALIGNMENTS

```

RESULT 1
A42435 leech antiplatelet protein precursor - Mexican leech
C:Species: Hemiptera officinalis (Mexican leech)
C>Date: 07-Apr-1994 #Sequence.Revision 07-Apr-1994 #text_change 07-May-1999
R:Accession: A42435; A42434
J: Biol. Chem. 267, 6899-6904, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42435; MUID:92202247; PMID:1551898
A:Accession: A42435
A:Molecule type: mRNA
A:Residues: 1-147 <KE>
A:Cross-references: GB:M81489
A>Note: the authors translated the codon AGC for residue 20 as Arg and GAA for residu
A>Note: the authors' translation is shown at residue 65
A>Note: sequence extracted from NCBI backbone (NCBI:91615) and modified
R:Connolly, T.M.; Jacobs, J.W.; Condra, C.
J: Biol. Chem. 267, 6893-6898, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42434; MUID:92202246; PMID:1551897
A:Accession: A42434
A:Molecule type: protein
A:Residues: 59-91;101-113;123-139 <CON>
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-147/Product: leech antiplatelet protein #status predicted <MAT>

Query Match 14.0%; Score 80.5; DB 2; Length 147;
Best Local Similarity 32.7%; Pred. No. 1.7;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

QY 5 DQWTFYANRRKTYDFDKSRKSSDLDCKKCTFKEY---CYIVEDTYNKKCCYNN 56
Db 66 DQWTKRPQKLPDNLTKTEFTSVDECRKCEESAVEPSCYLLQINTEETNCYNN 120
A55320
Immunophilin FKBP46 - fall armyworm
C:Species: Spodoptera frugiperda (fall armyworm)
C>Date: 06-Feb-1995 #Sequence.Revision 06-Feb-1995 #text_change 02-Sep-2000
R:Accession: A55320
R:Alammi, E.S.; Fernandes-Alnemri, T.; Pometenke, K.; Robertson, N.M.; Dudley, K.; D
J: Biol. Chem. 269, 30828-30834, 1994
A:Title: FKBP40, a novel S19 insect cell nuclear immunophilin that forms a protein-ki
A:Reference number: A55320; MUID:95074110; PMID:7527037
A:Accession: A55320
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-412 <AIN>
A:Cross-references: GB:U15038; NID:g595844; PID:g595845
C:Superfamily: yeast peptidylprolyl isomerase FPR3; BKBP-type peptidylprolyl isomerases

```


OY 67 KEVVDENFTENYLTDCGKAGNAAGTGDESDVE 102
 DB 428 KDEKEGEGDEGDGEDEGEDEG--DEGDEGDE 461

RESULT 7

T22054
 hypothetical protein F40G12.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
 C:Accession: T22054

R:McMurray, A.
 submitted to the EMBL Data Library, July 1996

A:Reference number: Z19507
 A:Accession: T22054
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-264 <WIL>
 A:Cross-references: EMBL:Z77661; PIDN:CA01183.1; GSPDB:GN00023; CESP:F40G12.5
 A:Experimental source: clone F40G12
 C:Genetics:
 A:Gene: CESP:F40G12.5
 A:Map position: 5
 A:Introns: 57/3: 158/3
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein F40G12.4

Query Match 13.2%; Score 75.5; DB 2; Length 264;
 Best Local Similarity 27.7%; Pred. No. 8.8;
 Matches 31; Conservative 19; Mismatches 33; Indels 29; Gaps 8;

OY 1 EERECCWTFYANKRTYTDGSKK--SSPLDECKTCFTECYIVFE-----DYVKE 52
 DB 29 ESEDEC-----FEKVFLIISGKHCCK-----DYDFARNLIQRREALTSGKE 72

OY 53 CYYNVVDGEELDOERF-VVDENFTE--NYLFDGCKDAGNAAGTDESDVD 101
 DB 73 CFLEIVK-ECEPBEKFKLIEINYSQVYLLTE-KPKDNGACTAPYFLEITE 122

RESULT 8

T16354
 hypothetical protein F4ZG9.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

R:Taich, A.
 submitted to the EMBL Data Library, March 1996

A:Description: The sequence of *C. elegans* cosmid F4ZG9.
 A:Reference number: Z18498
 A:Accession: T16354
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-491 <TAI>
 A:Cross-references: EMBL:U00051; NID:g1216305; PID:g1216310; PIDN:AAA91358.1; CESP:F4ZG9
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F4ZG9.1
 A:Introns: 37/3; 254/2; 364/2; 406/2

Query Match 13.2%; Score 75.5; DB 2; Length 491;
 Best Local Similarity 24.5%; Pred. No. 16;
 Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

OY 10 YANKRYDF--DKSEKSSDLDECKTCFTECYIVFEEDTVNKECYYNVVDGEELDOER 67
 DB 67 FTSAKLPDFLKERKFEADVAVAECLQKAF-----VDFDFIRAF-----ESKKEIK 112

OY 68 FVVDENFTENYLTDCGKAGNAAGTGDESDVD 101
 DB 113 DIGDEG-----KPKKAGGEADESEDEADRID 137

RESULT 9

T34510
 hypothetical protein ZK1290.10 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T34510

R:Taich, A.
 submitted to the EMBL Data Library, July 1995

A:Description: The sequence of *C. elegans* cosmid ZK1290.
 A:Reference number: Z21535
 A:Accession: T34510
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-360 <TAI>
 A:Cross-references: EMBL:U21308; PIDN:AAB9316.1; GSPDB:GN00020; CESP:ZK1290.10
 A:Experimental source: strain Bristol N2; clone ZK1290
 C:Genetics:
 A:Gene: CESP:ZK1290.10
 A:Map position: 2
 A:Introns: 23/1; 48/1; 158/3; 216/1; 264/1; 312/1; 331/3
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein ZK1290.10

Query Match 13.1%; Score 75; DB 2; Length 360;
 Best Local Similarity 33.3%; Pred. No. 13;
 Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

OY 46 EDYVNRKCYYNVVDGEELDOERFV-----DENFTENYLTDCGKAGN----AAGTG 94
 DB 224 EDYVGEF--NVDFDKLFDKIKYIRNDGSTNTETEQSTTEKTKYKGTIEVSVELGSG 280

OY 95 DESDEVED 103
 DB 281 DDEDEND 289

RESULT 10

T18402
 aspartate/aspartate rich protein - malaria parasite (*Plasmodium falciparum*) (fragment)
 C:Species: *Plasmodium falciparum*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T18402
 R:Barale, J.C.; Candelle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley, K.
 Infect. Immun. 65, 3003-3010, 1997
 A:Title: Plasmodium falciparum AARP1, a giant protein containing repeated motifs rich
 A:Reference number: Z18929; MUID:97378065; PMID:9234746
 A:Accession: T18402

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3844 <BAR>
 A:Cross-references: EMBL:Y08926; NID:e1154302; PID:e311435; PIDN:CAA70130.1
 C:Genetics:
 A:Gene: aarp1

Query Match 13.1%; Score 75; DB 2; Length 3844;
 Best Local Similarity 23.9%; Pred. No. 1.5e+02;
 Matches 21; Conservative 23; Mismatches 34; Indels 10; Gaps 3;

OY 1 EERECCWTFYANKRTYTD-----FDK-SFKSSDLDECKTCFTECYIVFEEDTVN 51
 DB 1726 EDEECQIKESYKSKSECNKKNKNIIFDSISVLKRNKIRKKNYMKKNKCYIYDDNNNK 1785

OY 52 ECVYNVVDGEELDOERFVVDENFTENYLT 79
 DB 1786 K-KKKKNVNEQKEFYVINKIFVINFI 1812

RESULT 11

T32731
 PAR interacting protein - rat

C:Species: *Rattus norvegicus* (Norway rat)
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32731
 R:Comte, P.A.; Ossipow, V.; Schibler, U.
 submitted to the EMBL Data Library, January 1997

A:Description: Isolation of PIP, a 160 kDa nucleolar protein that interacts with the actin
 A:Reference number: 221213
 A:Accession: T32731
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1277 <COM>
 A:Cross-references: EMBL:U83590; NID:g2253210; PID:g2253211; PIDN:AAB62878.1

Query Match 13.0%; Score 74.5; DB 2; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 53;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EREDCTWYANRKYTFD-KSPKSSDLECKKCFKTECYIVFEDTVNKECYNVVD 59
 DB 630 EDEED-----NVVDTDEKOLKHEGADSDS-----EDSKNSSDVDSDD 670
 QY 60 GEELDQE--KFVVDENFTENYLTDECKDAGNAAGTGDSEDE 99
 DB 671 GESEEDRDNDKVDPEFRQQLM---EVLQAGNALGSEEEDE 709

RESULT 12

D70465
 hypothetical protein aq.1923 - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: D70465
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; M0ID:98196666; PMID:9537320
 A:Accession: D70465
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-296 <AOF>
 A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07719.1; PID:g2984187; GB:AE00065
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: aq.1923
 C:Superfamily: Aquifex aeolicus hypothetical protein aq.1923

Query Match 12.9%; Score 74; DB 2; Length 296;
 Best Local Similarity 24.2%; Pred. No. 14;
 Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

QY 9 FYANRKT-----FDKSKSSDLECKKCFKTECYIVFEDTVNKECY 54
 DB 18 YFAYDKYMDKAKAEQVEYFLDKTLTKSG-----KGYIVYDKPIGGEIILKDY 67
 QY 55 YNVVDGEELDQEFVVDENFTENYLTDECKG 85
 DB 68 YRDNGEERFKIEELIIRK-----LSFECK 92

RESULT 13

B70236
 hypothetical protein BBH09 - Lyme disease spirochete plasmid H/1p28-3
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: B70236
 R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; M0ID:9805945; PMID:9403685
 A:Accession: B70236
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1278 <KLE>
 A:Cross-references: GB:AE000784; NID:g2690041; PIDN:AAC66000.1; PID:g2690056; TIGR:BBH09

A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 12.9%; Score 74; DB 2; Length 1278;
 Best Local Similarity 25.0%; Pred. No. 59;
 Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

QY 10 YANRKYDPEKSKSSDLECKKCFKTECYIVFEDTVNKECYNVV-----VDGEELD 64
 DB 181 YGEKRIEFD--FKRIKEE-----YKQEWFLFESYLIRKRYLKTNSIVSEKQIS 233
 QY 65 QKEFVVDENFT 76
 DB 234 KEKEIIOKTURE 245

RESULT 14

S17231
 dynein beta heavy chain, ciliary - sea urchin (Anthocidaris crassispina)
 N:Contains: dynein ATPase (EC 3.6.4.2)
 C:Species: Anthocidaris crassispina
 C:Date: 30-Sep-1991 #sequence_revision 02-May-1994 #text_change 19-Apr-2002
 C:Accession: S17231; PS0415
 R:Ogawa, K.
 Nature 352, 643-645, 1991
 A:Title: Four ATP-binding sites in the midregion of the beta heavy chain of dynein.
 A:Reference number: S17231; M0ID:91326104; PMID:1830928
 A:Accession: S17231
 A:Molecule type: mRNA
 A:Residues: 1-4466 <OGA>
 A:Cross-references: GB:D01021; NID:g217202; PIDN:BA00827.1; PID:g217203
 R:Ogawa, K.
 Proc. Jpn. Acad. B Phys. Biol. Sci. 67, 27-31, 1991
 A:Title: ATP-binding site in dynein beta-heavy chain: identification by molecular clo

A:Reference number: PS0415
 A:Accession: PS0415
 A:Molecule type: mRNA
 A:Residues: 764-1001, 'APG', 1005-2036, 'VPSSVET' <OG2>
 C:Superfamily: dynein heavy chain, ciliary
 C:Keywords: ATP; heterotrimer; hydrolase; microtubule binding; nucleotide binding;
 F:154-161/Region: nucleotide-binding motif A (P-loop)
 F:1857-1859/Region: nucleotide-binding motif A (P-loop)
 F:2133-2140/Region: nucleotide-binding motif A (P-loop)
 F:2460-2467/Region: nucleotide-binding motif A (P-loop)
 F:2805-2812/Region: nucleotide-binding motif A (P-loop)
 F:160/Binding site: ATP (Lys) #status predicted
 F:1858/Binding site: ATP (Lys) #status predicted
 F:2133/Binding site: ATP (Lys) #status predicted
 F:2466/Binding site: ATP (Lys) #status predicted
 F:2811/Binding site: ATP (Lys) #status predicted

Query Match 12.9%; Score 74; DB 1; Length 4466;
 Best Local Similarity 25.9%; Pred. No. 2.1e+02;
 Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

QY 1 EREDCTWYANRKY-----TDFKSKSSDLECKKCFKTECYIVFEDTVNKECYNVV-----CK-----KTC 35
 DB 460 EEEQCAKVFETPRYDGLDPTQCEPLEDEFEKRVFDDRLGSLTQCGPDCCGLEAA 519
 QY 36 FKTECY-----IVFEDTVNKECYNVV-----DGEELDQEFVVDEN 73
 DB 520 FKMLDCTGYLDRPVRINDF-----ECKYPLVLMVQELDQSKSEIYDEH 564

RESULT 15

T21428
 hypothetical protein p26F2.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21428
 R:Ainscough, R.
 submitted to the EMBL Data Library, November 1996

A: Reference number: 219420
 A: Accession: T21428
 A: Status: preliminary; translated from GB/EMBL/DDBJ
 A: Molecule type: DNA
 A: Residues: 1-643 <MIL>
 A: Cross-references: EMBL:Z81514; PIDN:CAB04192.1; GSPDB:GN00023; CESP:F26F2.6
 A: Experimental source: clone F26F2
 C: Genetics:
 A: Gene: CESP:F26F2.6
 A: Map position: 5
 A: Introns: 18/3; 116/3; 142/1; 265/3; 306/3; 373/3; 403/2; 425/2; 479/3; 544/3; 573/2; 6

Query Match 12.7%; Score 73; DB 2; Length 643;
 Best Local Similarity 26.6%; Pred. No. 37;
 Matches 29; Conservative 14; Mismatches 32; Indels 34; Gaps 6;

QY 7 WTFY-----ANRRYT---DFDKSFKK-----SDLDECKTKCFKTEYCYIVFED 47
 Db 155 WSFFRPYVDGCDANGSVTMVDYAKPGPETVYKPRDAQISNMDECMHCMEEECOIAYLD 214
 QY 48 TVNKEC-YYNVVVGEEILDQ-----KFVVDEN---FTENYLYTD 81
 Db 215 SDAKKCVVYSSDDGLTFMNKSSADSGKRLVIKMKVDETCLFTTPQLLD 263

Search completed: February 25, 2003, 02:33:20
 Job time : 16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:26:35 ; Search time 13 Seconds

(without alignments)
328.620 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EREDCWTFYANRKYTFDK.....GKDAAGNAGTGDESDDEVDED 103

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 85.5 | 14.9 | 147 | 1 | LAPP_HAEOR |
| 2 | 80.5 | 14.0 | 412 | 1 | FBK4_SPOFR |
| 3 | 78 | 13.6 | 797 | 1 | VG48_HSVSA |
| 4 | 75.5 | 13.2 | 419 | 1 | ICEC_MOUSE |
| 5 | 75.5 | 13.2 | 491 | 1 | P2C1_CAEEL |
| 6 | 75 | 13.1 | 360 | 1 | Y0RA_CAEEL |
| 7 | 74 | 12.9 | 296 | 1 | Y0Z3_MOUSE |
| 8 | 74 | 12.9 | 4466 | 1 | DYHC_ANTCR |
| 9 | 73 | 12.7 | 1405 | 1 | DPOA_SCHPO |
| 10 | 71 | 12.4 | 782 | 1 | YNY2_YEAST |
| 11 | 70.5 | 12.3 | 78 | 1 | MOH1_CANPG |
| 12 | 70.5 | 12.3 | 146 | 1 | PA21_MOUSE |
| 13 | 70.5 | 12.3 | 156 | 1 | CB20_HUMAN |
| 14 | 70.5 | 12.3 | 748 | 1 | Y875_METUA |
| 15 | 70 | 12.2 | 1058 | 1 | SI85_YEAST |
| 16 | 69.5 | 12.1 | 146 | 1 | PA21_PIG |
| 17 | 68.5 | 12.0 | 610 | 1 | CALG_HUMAN |
| 18 | 68 | 11.9 | 146 | 1 | PA21_MOUSE |
| 19 | 68 | 11.9 | 240 | 1 | NER_HY2D2 |
| 20 | 68 | 11.9 | 4466 | 1 | DYHC_TRIGR |
| 21 | 67.5 | 11.8 | 277 | 1 | PS12_ARATH |
| 22 | 67.5 | 11.8 | 602 | 1 | SPR8_YEAST |
| 23 | 67.5 | 11.8 | 602 | 1 | SPR8_YEAST |
| 24 | 67 | 11.7 | 146 | 1 | PA21_RAT |
| 25 | 67 | 11.7 | 327 | 1 | RREP_PIRYV |
| 26 | 67 | 11.7 | 650 | 1 | CC45_YEAST |
| 27 | 67 | 11.7 | 2410 | 1 | MOX1_SCHPO |
| 28 | 66.5 | 11.6 | 78 | 1 | MOX1_SCHPO |
| 29 | 66.5 | 11.6 | 125 | 1 | FABL_ANOPV |
| 30 | 66.5 | 11.6 | 132 | 1 | PA21_HORSE |
| 31 | 66.5 | 11.6 | 405 | 1 | LHX1_BRARE |
| 32 | 66.5 | 11.6 | 591 | 1 | CALX_RAT |
| 33 | 66.5 | 11.6 | 733 | 1 | YU08_YEAST |

| | | | | | | |
|----|------|------|------|---|------------|--------------------|
| 34 | 66.5 | 11.6 | 1033 | 1 | S190_YEAST | P36123 saccharomyc |
| 35 | 66 | 11.5 | 148 | 1 | CB20_XENIA | P52299 xenopus lae |
| 36 | 66 | 11.5 | 1647 | 1 | YDEL_SCHPO | O10435 schizosacch |
| 37 | 65.5 | 11.4 | 165 | 1 | ADFL_CAEEL | Q07750 caenorhabdi |
| 38 | 65.5 | 11.4 | 428 | 1 | RF4_KLULA | P09806 kluyveromyc |
| 39 | 65.5 | 11.4 | 455 | 1 | YNU6_YEAST | P40161 saccharomyc |
| 40 | 65.5 | 11.4 | 725 | 1 | RIR1_MYCTU | P50640 mycobacteri |
| 41 | 65.5 | 11.4 | 746 | 1 | GYP7_YEAST | P48365 saccharomyc |
| 42 | 65.5 | 11.4 | 793 | 1 | NETB_DROME | Q24568 drosophila |
| 43 | 65 | 11.3 | 1124 | 1 | TCF8_HUMAN | P37275 homo sapien |
| 44 | 65 | 11.3 | 1281 | 1 | TRPS_HUMAN | Q9uh17 homo sapien |
| 45 | 65 | 11.3 | 1616 | 1 | VIT1_CAEEL | P55155 caenorhabdi |

ALIGNMENTS

```

RESULT 1
LAPP_HAEOR
ID LAPP_HAEOR STANDARD: PRT: 147 AA.
AC 001747;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Anti-platelet protein precursor.
GN LAPP.
OS Haementeria officinalis (Mexican leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudindae; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Haementeria.
OX NCBI_TaxID=6410;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.
RC TISSUE=Salivary gland;
RX MEDLINE=92202247; PubMed=1551898;
RA Kellier P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;
RT "An inhibitor of collagen-stimulated platelet activation from the
RT salivary glands of the Haementeria officinalis leech. II. Cloning of
RT the cDNA and expression."
RL J. Biol. Chem. 267:6899-6904(1992)
CC -!- FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET
CC -!- AGGREGATION, DENSE GRANULE RELEASE AND SEROTONIN RELEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81489; AAA29194.1; -.
KW SIGNAL.
FT SIGNAL. 1 21 POTENTIAL.
FT CHAIN 22 147 ANTI-PLATELET PROTEIN.
SQ SEQUENCE 147 AA: 15908 MW: 75455113744AE42E CRC64;
Query Match 14.9%; Score 85.5; DB 1; Length 147;
Best local similarity 33.9%; Pred. No. 0.2;
Matches 19; Conservative 10; Mismatches 24; Indels 3; Gaps 1;
OY 4 EDCWTFYANRKYTFDKSKRSDDLECKKTCFTEY---CYIFEDTVNKECYN 56
DB 65 EDCWKRGRGKMLPDLNLTKEFTSVDECRKNCESNAVPSCTLIQINTETNECRN 120
RESULT 2
FKB4_SPOFR STANDARD: PRT: 412 AA.
ID FKB4_SPOFR
AC 026486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1999 (Rel. 38, Last annotation update)

```

DE 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
 DE isomerase) (PPIase) (EC 5.2.1.8).
 OS Spodoptera frugiperda (Fall armyworm).
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
 CC Ditrysia; Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
 ON NCBI_TaxID=7108;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=7527037;
 RX MEDLINE=95074110; PubMed=7527037;
 RA Alnemri E.S., Fernandes-Alnemri T., Pomeroy K., Robertson N.M.,
 RA Dudley K., Dudois G.C., Litwack G.;
 RT "FKBP6, a novel Sf9 insect cell nuclear immunophilin that forms a
 RT protein-kinase complex.";
 RL J. Biol. Chem. 269:30828-30834(1994).
 CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS
 CC DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
 CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
 CC PEPTIDE BONDS IN OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+
 CC AND ATP.
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U15038; AAA58962.1; -;
 DR HSSP: P20071; FRJ.
 DR InterPro: IPR001179; FKBP_PPIase.
 DR PIRam: PF00234; FKBP_1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
 DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
 KW Isomerase; Rotamase; Nuclear protein; DNA-binding; Phosphorylation.
 FT DOMAIN 91 112 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 120 145 LYS-RICH (BASIC).
 FT DOMAIN 152 216 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 219 302 LYS-RICH (BASIC).
 FT DOMAIN 324 412 PPIASE, FKBP-TYPE.
 FT SEQUENCE 412 AA; 45810 MW; P2A69159AEF4FE22 CRC64;
 SQ
 Query Match 14.0%; Score 80.5; DB 1; Length 412;
 Best Local Similarity 30.6%; Pred. No. 1.7;
 Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
 QY 11 ANKRYTDFDKSFKSS-----SDLECKKTKCTKCYIVFEDYVNRKCYVNVGSELDQ 65
 DB 133 ANKRAKPKDKAKGKNSAPAESDSDDD-----EDQLQK-----FLDGEDIDT 174
 QY 66 EKFDVDEFTENYLTDCGKDGAGNAAGTGDSEDEVD 103
 DB 175 DE--NDESFKN--TSAEGDSDSEDEDEDEDEDD 208
 RESULT 3
 VG48_HSVSA STANDARD: PRT; 797 AA.
 AC 001033;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Hypothetical gene 48 protein.
 GN 48 OR EDL5.
 OS Herpesvirus saimiri (strain 11).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Gammaherpesvirinae; Rhadinovirus.
 CC NCBI_TaxID=10383;
 CC

RN [1]
 RP SEQUENCE FROM N.A. PubMed=1321287;
 RX MEDLINE=92333688; PubMed=1321287;
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisinger B.,
 RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.W.;
 RT "Primary structure of the herpesvirus saimiri genome.";
 RL J. Virol. 66:5047-5058(1992).
 CC -1- SIMILARITY: TO EBV BRRF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X64346; CAA45671.1; -;
 DR PIR: A36811; A36811.
 KW Hypothetical protein.
 SQ SEQUENCE 797 AA; 88667 MW; 9E294234AD850E23 CRC64;
 SQ
 Query Match 13.6%; Score 78; DB 1; Length 797;
 Best Local Similarity 32.3%; Pred. No. 5.7;
 Matches 31; Conservative 12; Mismatches 33; Indels 20; Gaps 5;
 QY 11 ANKRYTDFDKSFKSSDLECKKTKCTKCYIVFEDYVNRKCYN---VVDGELEDE 66
 DB 382 ANKR--EYKTIIDSDRDRODRDDEYLE-----NEE--VNRDEEDEGEDEDE 427
 QY 67 KFDVDEFTENYLTDCGKDGAGNAAGTGDSEDEVD 102
 DB 428 KDKREGDEBDGDEGEDEGEDEGE--DEGDEGEDEDE 461
 RESULT 4
 ICEC_MOUSE STANDARD: PRT; 419 AA.
 AC 008736;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Caspase-12 precursor (EC 3.4.22.-).
 GN CASP12.
 OS Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/An.
 RX MEDLINE=97190206; PubMed=9038361;
 RA van de Craen M., Vandanaebele P., Declercq W., van den Brande I.,
 RA van Looy G., Molemans F., Schotte P., van Criekinge W., Beyaert R.,
 RA Fiers W.;
 RT "Characterization of seven murine caspase family members.";
 RL FEBS Lett. 403:61-69(1997).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN SKELETAL MUSCLE AND LUNG.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
 CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

[illegible]

OY 95 DESDEVED 103
: : : : :
Db 281 DDEENDD 289

RESULT 7

ID YJ23_AOUAE STANDARD; PRT; 296 AA.
AC 067752;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein A01923 precursor.
GN A01923.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Aujay M., Huber R.,
RA Reldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: A000763; AAC07719.1; -
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 296 HYPOTHETICAL PROTEIN A01923.
SQ SEQUENCE 296 AA; 34566 MW; 9C88F4456A9449C CRC64;

Query Match 12.9%; Score 74; DB 1; Length 296;
Best Local Similarity 24.2%; Pred. No. 4.8;
Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

OY 9 FANRKYD-----FDKSEKSSDLDECKKTCFTECYIVE---DTVNECY 54
Db 18 YAYDKYMDNKAKEQVEYFLDKTLRKS-----KSKYKVDKPIGGLITIKDY 67
OY 55 YVVDGEELDEKPFVDENFTENLTDOCEK 85
Db 68 YDRNGEERKEIEIIEK-----LSETECK 92

RESULT 8

ID DYC_ANTCR STANDARD; PRT; 4466 AA.
AC P39057;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dyein beta chain, ciliary.
OS Anthocidaris crassispina (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Anthocidaris; Echinoidea; Echinodermata; Echinoida; Echinometridae;
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91326104; PubMed=1830928;
RA Ogawa K.;

RT "Four ATP-binding sites in the midregion of the beta heavy chain of
RT dynein.";
RL Nature 352:643-645(1991).

CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYEIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: D01021; BAA0827.1; -
DR PIR: S17231; S17231.
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy_1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
FT COILED COIL.
FT DOMAIN 482 502 COILED COIL (POTENTIAL).
FT DOMAIN 627 643 COILED COIL (POTENTIAL).
FT DOMAIN 734 805 COILED COIL (POTENTIAL).
FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
FT DOMAIN 1950 1978 COILED COIL (POTENTIAL).
FT DOMAIN 3033 3334 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
FT NP_BIND 154 161 ATP (POTENTIAL).
FT NP_BIND 1852 1859 ATP (POTENTIAL).
FT NP_BIND 2133 2140 ATP (POTENTIAL).
FT NP_BIND 2460 2467 ATP (POTENTIAL).
FT NP_BIND 2805 2812 ATP (POTENTIAL).
SQ SEQUENCE 4466 AA; 511772 MW; C465C5C6C6D325D CRC64;

Query Match 12.9%; Score 74; DB 1; Length 4466;
Best Local Similarity 25.9%; Pred. No. 81;
Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

OY 1 EERECWTFYANKR-----TDFKSEKSSDLDE-----CK-----KTC 35
Db 460 EEFQCARVETEPYDGDLPYQEFLEDEFEKRVFLDRRLGSLICQGFDDCGLMA 519
OY 36 FTERCY-----IVFEDYFNKCYNNV---DGEELDEKRVVDEN 73
Db 520 FRMLDCYGPLLDRVYINDF---ECKIPVLMYLDQLDQSKETIDEH 564

RESULT 9

ID DPOA_SCHPO STANDARD; PRT; 1405 AA.
AC P28040;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase alpha catalytic subunit (BC 2.7.7.7) (DNA polymerase
DE 1).
GN POL1 OR SM17 OR SPAC3H5.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetales;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91238692; PubMed=2034212;
RA Damagnez V., Tillitt J., de Recondo A.-M., Baldacl G.;

DT 15-JUL-1998 (Rel. 36, last sequence update)
 DE 15-JUN-2002 (Rel. 41, last annotation update)
 DE Mandibular organ-inhibiting hormone-1 (MOIH-1).
 OS Cancer pagurus (Rock crab).
 OC Eucarysta: Metazoa: Arthropoda: Mandibulata: Pancrustacea: Crustacea:
 OC Malacostraca: Eumalacostraca: Eucarida: Decapoda: Pleocyemata;
 OC Brachyura: Eubrachyura: Cancroidea: Cancridae; Cancer.
 NCBI_TaxID=6755;
 RN [1]
 RP SEQUENCE.
 RC TISSUE-Sinus gland;
 RX MEDLINE=96278659; PubMed=8662685;
 RA Mainwright G., Webster S.G., Wilkinson M.C., Chung J.S., Rees H.H.,
 RT "Structure and significance of mandibular organ-inhibiting hormone in
 RT the crab, Cancer pagurus. Involvement in multihormonal regulation of
 RT growth and reproduction.";
 RL J. Biol. Chem. 271:12749-12754 (1996).
 CC -1- FUNCTION: REPRESENTS THE SYNTHESIS OF METHYL FARNESATE, THE
 CC PRECURSOR OF INSECT JUVENILE HORMONE III IN THE MANDIBULAR ORGAN.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE MEDULLA TERMINALIS X-ORGAN IN
 CC THE EUSTALKS AND TRANSPORTED TO THE SINUS GLAND WHERE IT IS
 CC STORED AND RELEASED.
 CC -1- SIMILARITY: BELONGS TO THE ANTHROPOD CHH/MTH/GH/VIH FAMILY OF
 CC HORMONES.
 DR InterPro: IPR001166; CHH_MTH_GH.
 DR Pfam: PF01147; Crust_neurohorm; 1.
 DR PRINTS: PR00550; HYPERGLICEMIC.
 DR PROSITE: PS01250; CHH_MTH_GH; 1.
 DR KW Neuropeptide; Hormone.
 FT DISULFID 7 44 BY SIMILARITY.
 FT DISULFID 24 40 BY SIMILARITY.
 FT DISULFID 27 53 BY SIMILARITY.
 SQ SEQUENCE 78 AA; 9241 MW; 8DB338A39058A62D CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 78;
 Best Local Similarity 33.9%; Pred. No. 2.5;
 Matches 20; Conservative 5; Mismatches 23; Indels 11; Gaps 3;
 Oy 5 DCMFTYNNRRY-----TDFKSPKSSDDECKKCE-KTEV-CYTFEDTVNKE 52
 Db 6 DCONFTGNRAMYKRVKWDICKCANIFRKDGLNNRSCNCFYTFETWCIDATENTRNKE 64
 RESULT 12
 PA21_CANFA STANDARD; PRT; 146 AA.
 AC P06596;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-1988 (Rel. 06, Last sequence update)
 DE Phospholipase A2 precursor (EC 3.1.1.4) (phosphatidylcholine 2-
 DE acylhydrolase) (Group IB phospholipase A2).
 GN PLA2G1B.
 OS Canis familiaris (Dog).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 OC Mammalia: Eutheria: Carnivora: Fissipedia: Canidae; Canis.
 NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86223862; PubMed=3754861;
 RA Ohara O., Tanaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
 RT Teraoka H., Okamoto M.;
 RT "Dog and rat pancreatic phospholipases A2: complete amino acid
 RT sequences deduced from complementary DNAs.";
 RL J. Biochem. 99:733-739 (1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90306027; PubMed=2142076;
 RA Kerfelec B., Latorge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
 RT "Isolation and sequence of the canine pancreatic phospholipase A2
 RT gene.";
 RL Eur. J. Biochem. 190:299-304 (1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87175472; PubMed=3562437;
 RA Kerfelec B., Latorge K.S., Puigserver A., Scheele G.A.;
 RT "Primary structures of canine pancreatic lipase and phospholipase A2
 RT messenger RNAs.";
 RL Pancreas 1:430-437 (1986).
 CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the
 CC 2-acyl groups in 3-sn-phosphoglycerides.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: D00035; BAA00023.1;
 DR EMBL: M35301; AAA30883.1;
 DR PIR: A24392; PSDG.
 DR PIR: S11316; S11316.
 DR HSSP: P00592; 4P2P.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF00068; Phoslip; 1.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR PRODOM: P000303; PhospholipaseA2.
 DR SMART: SM00085; PA2c; 1.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 DR KW Hydrolyase; Lipid degradation; Calcium; Pancreas; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 15
 FT CHAIN 23 146
 FT ACT_SITE 23 146
 FT ACT_SITE 70 70
 FT ACT_SITE 121 121
 FT DISULFID 33 99 BY SIMILARITY.
 FT DISULFID 49 146 BY SIMILARITY.
 FT DISULFID 51 146 BY SIMILARITY.
 FT DISULFID 66 127 BY SIMILARITY.
 FT DISULFID 73 120 BY SIMILARITY.
 FT DISULFID 83 113 BY SIMILARITY.
 FT DISULFID 106 118 BY SIMILARITY.
 FT CA_BIND 50 50 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 52 52 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 54 54 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 71 71 VIA CARBONYL OXYGEN (BY SIMILARITY).
 SQ SEQUENCE 146 AA; 16235 MW; F6258ED9527F3692 CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 146;
 Best Local Similarity 25.4%; Pred. No. 4.9;
 Matches 15; Conservative 11; Mismatches 18; Indels 15; Gaps 1;
 Oy 28 LDECKTKCFKTEFYCYVEEDVNNKECYNVWDGELDQEFVVDNFTNYLTDEGDK 86
 Db 60 VDELDKCOCTHDHC-----SEAKILDSCKFLDNDNYTYKYSYSGSGS 103
 RESULT 13
 CB20_HUMAN STANDARD; PRT; 156 AA.
 AC P52298; Q14924;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 20 kDa nuclear cap binding protein (NCBP 20 kDa subunit) (CBP20) (NCBP
 DE interacting protein 1) (NIP1).
 GN NCBP2 OR CBP20.
 OS Homo sapiens (human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 9-25 AND 113-145.
RX MEDLINE=95379956; PubMed=7651522;
RA Izurzaide E., Lewis J., Gambrell C., Jarmolowski A.,
RA McGuigan C., Malta A.W.;
RT "A cap-binding protein complex mediating U snRNA export.";
RL Nature 376:709-712(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=96017765; PubMed=7478990;
RA Kataoka N., Ohno M., Mada I., Shimura Y.;
RT "Identification of the factors that interact with NCBP, an 80 kDa
RT nuclear cap binding protein.";
RL Nucleic Acids Res. 23:3638-3641(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN MEDIATING U SNRNA EXPORT FROM THE NUCLEUS.
CC BINDS TO 5/CAPPED MENA.
CC -1- SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
CC OF CBP80 AND CBP20.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X84157; GAA58962.1; -
DR EMBL: D59253; BAA09599.1; -
DR EMBL: BC001255; AAH01255.1; -
DR HSSP: P11940; ICVJ.
DR GENE: HGNC:7659; NCBP2.
DR MIM: 605133; -
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 1.
DR SMART: SMD0360; RRM; 1.
DR PROSITE: PS0102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding.
FT DOMAIN 40 118 RNA-BINDING (RRM).
FT CONFLICT 97 97 A -> S (IN REF. 2).
SQ SEQUENCE 156 AA; 1801 MW; B6C94F3182A2CC3D CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 156;
Best Local Similarity 21.8%; Pred. No. 5.2;
Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;

QY 1 EERE-----DCWTFYANKRYTDFPKS-----FKKSSD-----LDCKKTCFTECYTV 44
DB 30 EDEKLLKSKCTLYVGMISYTTTBEQYELFSKSGDIKIMLDKMKRTA--CGFCFWE 87
QY 45 FEDTVNKECYNVVGEELDQEKFFVDENFTENYITDCEGKDGAGNAAGTGDSEDEVD 103
DB 88 YYSRADAENMRYINGTRLDLR--IIRTDWDAFK--EGROYGKRSRGGOYRDEYRQD 141

```

```

RESULT 14
Y875_METUA STANDARD; PRT; 748 AA.
AC 058285;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M0875.
GN M0875.
OS Methanococcus jannaschli.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA Scott J.L., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschli.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67531; AAB98882.1; -
DR TIGR: M0875; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 748 AA; 90731 MW; 5515ED3B541562NA CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 748;
Best Local Similarity 24.4%; Pred. No. 27;
Matches 29; Conservative 16; Mismatches 39; Indels 35; Gaps 6;

QY 1 EEREDCWTFY-----ANKRYTDFPKSFKSSDDECKTCFK-TECYIVFED--- 47
DB 144 KEFSIDWVYFDIKIYKYKLSIKRHKFGNDKAIELTE--KQYLAESYNNKDKNRY 200
QY 48 -----TVNKECYNVVGEELDQEKFFVDENFTENYITDCEGKDGAGNAAGTGDSEDEVD 102
DB 201 KKAELTFNHFYNNIM-----AQKFESERKFE-----AAEYKKSGDITKEIDE 244

```

```

RESULT 15
S185_YEAST STANDARD; PRT; 1058 AA.
AC P40856;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE S184-associated protein SAPI85.
GN SAPI85 OR YJL098W OR J0840
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96220458; PubMed=8649382;
RA Luke M.M., della Seta F., di Como C.J., Sugimoto H., Kobayashi R.,
RA Arndt K.T.;
RT "The SAP, a new family of proteins, associate and function positively
RT with the S184 phosphatase.";
RL Mol. Cell. Biol. 16:2744-2755(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;

```

```

RX MEDLINE-95176706; PubMed-7871887;
RA Miosga T., Boles E., Schaaf-Gerstenschlaeger I., Schmitt S.,
RA Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
RT Saccharomyces cerevisiae chromosome X including the BCK1 gene.";
RL Yeast 10:1481-1488(1994).
RM [3]
RP SEQUENCE OF 1-853 FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE-96090136; PubMed-7483851;
RA Rasmussen S.W.;
RT "A 37.5 kb region of yeast chromosome X includes the SWE1, MER2, GSH1
RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar
RT to the DAL80 gene, and a tRNA(Arg).";
RL Yeast 11:873-883(1995).
CC -! FUNCTION: ASSOCIATES WITH THE SIT4 PHOSPHATASE IN A CELL CYCLE
CC DEPENDENT MANNER. MAY BE DIRECTLY OR INDIRECTLY INVOLVED IN
CC SIT4-DEPENDENT FUNCTIONS IN BUDDING AND IN NORMAL G1 CYCLIN
CC EXPRESSION.
CC -! PTM: HYPERPHOSPHORYLATED IN THE ABSENCE OF SIT-4.
CC -! SIMILARITY: BELONGS TO THE SAPS FAMILY.
CC -----
CC this SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77923; CAA54882.1; -
DR EMBL: X85021; CAA59396.1; -
DR EMBL: Z49373; CAA89392.1; -
DR SGD: S0003634; SAPI85.
KW Phosphorylation; Cell cycle.
SQ SEQUENCE 1058 AA; 121402 MW; DD44DD53DD87438 CRC64;

Query Match 12.2%; Score 70; DB 1; Length 1058;
Best Local Similarity 20.3%; Pred. No. 43;
Matches 25; Conservative 27; Mismatches 45; Indels 26; Gaps 5;

QY 2 ERECCWTFYANRRKYTFDFKSFKKSDLDDECKTKCTFTEY-----CYIVE-- 46
DB 427 DRDPIYIGLYVKMESEHMAFNKI--LTEKKIPLLOTSTYGTIEPLGFERRFKICELIAEL 484
QY 47 -----DFVNRKCYVNVVGEELDEKEKYVDENFTE-NYLDOGKXGAGNAGTGDSDEY 100
DB 485 HCSNMTLNPSAYDIVREDAERERLEFNSQNYVDSNRSSEKLENEDN---TGDADDEV 541
QY 101 DED 103
DB 542 EDD 544

```

Search completed: February 25, 2003, 02:31:32
 Job time : 15 secs

| | | | | | | |
|----|------|------|------|----|--------|---------------------|
| 17 | 73.5 | 12.8 | 1154 | 5 | Q9V659 | Q9V659 drosophila |
| 18 | 73 | 12.7 | 643 | 5 | Q9XV53 | Q9XV53 caenorhabd1 |
| 19 | 73 | 12.7 | 2543 | 5 | P90649 | P90649 paramedium |
| 20 | 72.5 | 12.7 | 564 | 5 | O01627 | O01627 caenorhabd1 |
| 21 | 72.5 | 12.7 | 2533 | 5 | Q27183 | Q27183 paramedium |
| 22 | 72.5 | 12.7 | 2533 | 5 | P90589 | P90589 paramedium |
| 23 | 72 | 12.6 | 1303 | 10 | Q9SKS0 | Q9SKS0 arabidopsis1 |
| 24 | 71.5 | 12.5 | 320 | 5 | Q96214 | Q96214 plasmodium |
| 25 | 71.5 | 12.5 | 540 | 16 | Q8XHG9 | Q8XHG9 clostridium |
| 26 | 71.5 | 12.5 | 594 | 16 | Q9P661 | Q9P661 campylobact |
| 27 | 71 | 12.4 | 551 | 4 | Q9NX93 | Q9NX93 homo sapien |
| 28 | 71 | 12.4 | 589 | 4 | Q9HA35 | Q9HA35 homo sapien |
| 29 | 71 | 12.4 | 851 | 4 | Q9H501 | Q9H501 homo sapien |
| 30 | 70.5 | 12.3 | 112 | 5 | Q9U5Y6 | Q9U5Y6 cancer pagu |
| 31 | 70.5 | 12.3 | 256 | 10 | Q9CA59 | Q9CA59 arabidopsis1 |
| 32 | 70.5 | 12.3 | 296 | 11 | Q9D5G6 | Q9D5G6 mus musculus |
| 33 | 70.5 | 12.3 | 367 | 5 | Q9V1R9 | Q9V1R9 drosophila |
| 34 | 70.5 | 12.3 | 600 | 16 | Q8RH56 | Q8RH56 fusobacteri1 |
| 35 | 70.5 | 12.3 | 786 | 10 | Q9C7E7 | Q9C7E7 arabidopsis1 |
| 36 | 70.5 | 12.3 | 2472 | 5 | Q8FTM5 | Q8FTM5 dictyosteli |
| 37 | 70 | 12.2 | 292 | 12 | Q65239 | Q65239 african sw1 |
| 38 | 70 | 12.2 | 328 | 5 | Q22144 | Q22144 caenorhabd1 |
| 39 | 70 | 12.2 | 521 | 5 | Q94718 | Q94718 paramedium |
| 40 | 70 | 12.2 | 669 | 5 | Q8SX58 | Q8SX58 drosophila |
| 41 | 70 | 12.2 | 703 | 10 | Q9LKR9 | Q9LKR9 arabidopsis1 |
| 42 | 70 | 12.2 | 793 | 10 | Q9SPU5 | Q9SPU5 zea mays (m |
| 43 | 70 | 12.2 | 1167 | 16 | Q9ZLG7 | Q9ZLG7 helicobacter |
| 44 | 70 | 12.2 | 1255 | 16 | Q97I16 | Q97I16 clostridium |
| 45 | 70 | 12.2 | 1481 | 5 | Q9NEX0 | Q9NEX0 caenorhabd1 |

RESULT 1

| | | | | |
|----|--|--------------|------|---------|
| AC | 08XJ41: | PRELIMINARY; | PRT; | 116 AA. |
| ID | 08XJ41: | | | |
| DT | 01-MAR-2002 (TREMBLel. 20, Created) | | | |
| DT | 01-MAR-2002 (TREMBLel. 20, Last sequence update) | | | |
| DT | 01-JUN-2002 (TREMBLel. 21, Last annotation update) | | | |
| DE | Hypothetical protein CPE1920. | | | |
| GN | CPE1920. | | | |
| OS | Clostridium perfringens. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia; | | | |
| OC | Clostridiales; Clostridaceae; Clostridium. | | | |
| OX | NCBI_Taxid=1502; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=13 / TYPE A; | | | |
| RX | PubMed=11792842; | | | |
| RA | Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., | | | |
| RA | Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H., | | | |
| RT | "Complete genome sequence of Clostridium perfringens, an anaerobic | | | |
| RT | "flesh-eater."; | | | |
| RC | Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002). | | | |
| RL | EMBL; AP063192; BAB81626.1; -. | | | |
| DR | InterPro: IPR001440; TPR. | | | |
| DR | Pfam: PF00515; TPR; 3. | | | |
| DR | SMART: SM00028; TPR; 3. | | | |
| KW | Hypothetical protein; Complete proteome. | | | |
| QV | SEQUENCE 116 AA; 13590 MW; 9D6BSD847B3A7312 CRC64; | | | |

| query match | 13.78; | Score 78.5; | DB 16; | Length 116; |
|-------------|--------|-------------|--------|-------------|
|-------------|--------|-------------|--------|-------------|

| | | | | | | | | | |
|---------|-----|--------------|-----|------------|-----|--------|-----|------|---|
| Matches | 22; | Conservative | 11; | Mismatches | 27; | Indels | 37; | Gaps | 2 |
|---------|-----|--------------|-----|------------|-----|--------|-----|------|---|

QY 10 YANRKYTDFDKSEFKSSDLDECK-----KTC 35

Db 11 YNTRNSDAINYYKKALNDECKCHSYNAGVCYIKLKQYEKAIEMTKALELYQDSKYF 70

QY 36 FKTEYCYIVFEDTVNKECYNYN --- VDGELDQEKV 65

OC Bacillaceae: Bacillus.
 NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.,
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.",
 RT Nucleic Acids Res. 28:4317-4331(2000).
 CC -1 SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DR EMBL: AF001517, BAB06889.1;
 DR HSSP: 056310, 1B30.
 DR InterPro: IPR003584; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR002345; Chem.
 DR InterPro: IPR004105; H-kinase_dtm.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR002570; Hpt.
 DR Pfam: PF01584; Chem; 1.
 DR Pfam: PF02895; H-kinase_dtm; 1.
 DR Pfam: PF02318; HATPase_c; 1.
 DR Pfam: PF01627; Hpt; 1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR PRODOM: PD003142; Hpt; 1.
 DR SMART: SM00260; Chem; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00073; Hpt; 1.
 DR Kinase: Phosphorylation; Sensory transduction; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 682 AA; 75272 MW; 46B6B1E5BEF7276 CRC64;

Query Match 13.6%; Score 78; DB 16; Length 682;
 Best Local Similarity 32.6%; Pred. No. 15;
 Matches 26; Conservative 13; Mismatches 31; Indels 14; Gaps 5;

QY 10 YANRKYTDKFSFKSSD-----LDECKTKCTKTECYVFE--DTVNEKCYNNYVG 60
 DB 156 YDEFEMVTLDSFEGYAYOIEVTLDE--KTLTKARVFEVLEBOVG-EVTKSPSPA 212
 QY 61 EELDEKFEVDENFENYLTDCDGD 86
 DB 213 EELDEKFE--DERFLVTLITVNDGE 236

RESULT 6

097239 PRELIMINARY; PRT; 3933 AA.

AC 097239
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 467.9 kDa protein.
 GN PFC0245C, MAL3P2.18.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Chutcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
 RA Gentles S., Gilliam R., Hamlin N., Harris D., Holtroyd S., Hornsby T.,
 RA Hurrell K., Jagels K., Jaesal B., Kyes S., McLean J., Moule S.,
 RA Murrell K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skellon J., Squares R., Squares S., Sulston J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
 RT falciparum".
 RL Nature 400:532-538(1999).

DR EMBL: AL034558; CAB39005.1;
 DR InterPro: IPR002048; EF-hand.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 3933 AA; 467876 MW; 5144A4604EE36933 CRC64;

Query Match 13.6%; Score 78; DB 5; Length 3933;
 Best Local Similarity 25.0%; Pred. No. 1e+02;
 Matches 23; Conservative 21; Mismatches 24; Indels 24; Gaps 5;

QY 36 FKTEYC-YIVF-----EDTVNKECY--NVYDGEILD-QEKF-----VVD 71
 DB 1139 FETNNCEFIIFLFYVHLLIEKEKLNKSCFELKNIINNCCEMKNKFEVFCFLHLIR 1198
 QY 72 ENFENYLTDCGKDGAGNAAGTGDSEDEVDD 103
 DB 1199 INFNKSLLKNYKIKRAGDITDDDDDDDDDD 1230

RESULT 7

09VFEH8 PRELIMINARY; PRT; 368 AA.

AC 09VFEH8
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG14853 protein.
 GN CG14853.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne C.R., McKlos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt R.A., Nelson C.R., Pfaffner B.D.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Boulter J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fouts R.C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison U.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

[illegible]

| | | | |
|-----------|--|---|---------------|
| D6 | 73 | CELEIYK-BECEPEKFLIENYQIVLTLLE-KPKNGACGAPYPOLEIE | 122 |
| RESULT 11 | | | |
| ID | 094648 | PRELIMINARY; | PRT; 3844 AA. |
| AC | 094648; | | |
| DT | 01-FEB-1997 (TRMBLrel. 02, Created) | | |
| DT | 01-JUL-1997 (TRMBLrel. 04, Last sequence update) | | |
| DT | 01-DEC-2001 (TRMBLrel. 19, Last annotation update) | | |
| DE | AAPI protein (Fragment). | | |
| GN | AAPI. | | |
| OS | Plasmodium falciparum (isolate 3D7). | | |
| OC | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | |
| OX | NCBI_TaxID=36329; | | |
| RRN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | SPRAIN=3D7; | | |
| RX | MEDLINE=97378065; PubMed=9234746; | | |
| RA | Barale U.C., Candelie D., Attal-Bonnefoy G., Dehoux P., Bonnefoy S., Ridley R., Pereira da Silva L., Langsley G.; diagnosing protein containing repeated | | |

| Query Match | Score | DB | Length |
|--|-------|----|--------|
| FT motifs rich in asparagine and aspartate residues, is associated with the infected erythrocyte membrane."; | 75 | DB | 3844 |
| RT Infect. Immun. 65:3003-3010(1997). | 75 | DB | 3844 |
| EMBL: Y08925: CAAT70130.1; - | 75 | DB | 3844 |
| DR InterPro: IPR001841; ZnF_RING. | 75 | DB | 3844 |
| DR SMART, SM00184; RING: 1. | 75 | DB | 3844 |
| FT NON-TER | 75 | DB | 3844 |
| FT NON-TER | 75 | DB | 3844 |
| SEQ SEQUENCE | 75 | DB | 3844 |

Maiores

[illegible]

AC Osr1w01
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE GB|AAE22924.1 (Hypothetical 71.0 kba protein).
GN A75G39780.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosida
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
FX MEDLINE=99087489; PubMed=9672454;
ME

RA Nakamura I., Saito C., Yanagida M., Yanagida S.,
RA Tabata S.,
RA "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT "Structural features of the regions of 1,013,767 bp covered by sixteen
RT sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned PI and TAC clones.",
RT DNA Res. 5:297-308(1998).
RL [2]
RM
RM
RM
RP
RP
RP
RP Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Tortum M., Wu H.C.,

RA Yamamura Y., Yu G., Yu S., Bowser L., Carinci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlén-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full length cDNA of gene At5g39780 (GI:15242471)."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016876; BAB11383.1; -
 DR EMBL; AY070102; AAL49839.1; -
 KW Hypothetical protein
 SQ SEQUENCE 606 AA; 71043 MW; 5F02E244E59FC15D CRC64;

Query Match 13.0%; Score 74.5; DB 10; Length 606;
 Best Local Similarity 27.7%; Pred. No. 29;
 Matches 31; Conservative 19; Mismatches 37; Indels 25; Gaps 7;
 QY 1 EEREPCW--TFYANRKYTFD---KSKKSSDLDECKKCTCYCYIVFEDTVNKECY 55
 DB 103 EKTEDLGVSYPHGDKKTKEDYSVSFKK-----KTRIFLFE-----EDFLESDS-- 147
 QY 56 NVVDGEEL---DOEKVVDENFTENYLTDCGKDGAGNAGTGDSEDEVD 103
 DB 148 DFDVDSQTFSTNDEDFLSDPFAETSLKKGKNSDN--SGSGSDESEEE 198

RESULT 13
 ID 035821 PRELIMINARY; PRT; 1277 AA.
 AC 035821;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PAR interacting protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91084854; PubMed=2261643;
 RX Martin J., Schibler U.;
 RT "Expression of the liver-enriched transcriptional activator protein
 RT DBP follows a stringent circadian rhythm.";
 RL Cell 63:1257-1266 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Conte P.A., Ossipow V., Schibler U.;
 RT "Isolation of Plp, a 160 kDa nuclear protein that interacts with the
 RT activation domain of PAR transcription factors.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U83590; AAB62878.1; -
 SO SEQUENCE 1277 AA; 144675 MW; 43140A2973178792 CRC64;

Query Match 13.0%; Score 74.5; DB 11; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 66;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;
 QY 1 EEREPCWTFYANRKYTFD--KSKKSSDLDECKKCTCYCYIVFEDTVNKECY 59
 DB 630 EDEED-----NVVDTDTDEKQLKHGSDADS-----EDSKNSSEVDSDSD 670
 QY 60 GEELDQ--KVVVDENFTENYLTDCGKDGAGNAGTGDSEDE 99
 DB 671 GESEEDRDKVDYDPRGQOLM--EVLQAGNALGSEEEEEE 709

RESULT 14
 ID 09FK41 PRELIMINARY; PRT; 596 AA.
 AC 09FK41;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similarity to unknown protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RX MEDLINE=98403884; PubMed=9734815;
 RA Krcani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Physically assigned pi and TAC clones.";
 RL DNA Res. 5:203-216 (1998).
 DR EMBL; AB012248; BAB09450.1; -
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF01344; Kelch; 6.
 SO SEQUENCE 596 AA; 68483 MW; 8665D3CAB873CB0D CRC64;

Query Match 12.9%; Score 74; DB 10; Length 596;
 Best Local Similarity 24.2%; Pred. No. 32;
 Matches 29; Conservative 23; Mismatches 40; Indels 28; Gaps 5;
 QY 12 NRRYTFDNSEK-----KSSDLDECKKCTCYCYIVFEDTVN-----NKEC 53
 DB 339 NRRMSDKSKALQCAKTRLPQVSVSEYVKKPGCRINSNVGKDTLYGGMEIKKEV 398
 QY 54 YVN--VVGGEELDQKPFVVDENFTENYLTDCG-----KDAGNAGTGDSEDEVD 103
 DB 399 TLDDLVSLSMKLDEWKCIIPTEWEVSDSDGDEDDDESDSESGNSESDDDDD 458

RESULT 15
 ID 050667 PRELIMINARY; PRT; 1278 AA.
 AC 050667;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein BBH09.
 GN BBH09.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathiera R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson S., Hanson M.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weiman J.,
 RA Uterback T., Watney L., McDonald L., Attlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586 (1997).
 DR EMBL; AE000784; AAC66000.1; -
 DR TIGR; BBH09; -
 DR InterPro; IPR002296; N12N6_mtfase.
 DR InterPro; IPR002052; N6_mtfase.
 DR PRINTS; PR00507; N12N6_mtfase.
 DR PROSITE; PS00092; N6_mtfase; UNKNOWN 1.
 KW Hypothetical protein; Plasmid; Complete proteome.
 SO SEQUENCE 1278 AA; 150871 MW; 8A0B22DF166565C0 CRC64;

Query Match 12.9%; Score 74; DB 16; Length 1278;

Wed Feb 26 15:59:07 2003

us-09-936-737a-2.rspt

Page 7

Best Local Similarity 25.0%; Pred. NO. 74;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

```

QY      10 YANRKYTDDEKSEKSSDLDECKKTCIKTEKYIVFEETVNKEGYNV-----VDGEELD 64
        | ::|| | : : | : : || | : : | : : | : : | : : | : : | : : |
Db      181 YGKRYIEFD-FSKIKEEE---YKEQEWELFSLYLRKERYLKTSNISIVEKEQIS 233

```

```
QY      65 QEKFEVDENFTE 76
        :||::: |
Db      234 KEKEIIQKTLRE 245
```

Search completed: February 25, 2003, 02:32:52
Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:33:51; Search time 1946 Seconds

(without alignments)
857.212 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EBERDCWTFYANKRYTDFDK.....GKDAAGNAAGTDESDEVED 103

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US09936737/runat_14022003_100733_2299/app_query.fasta_1.263
-DB=Est -QFMT=fastap -SUFFIX=st -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIS=45
-LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09936737@cgn2.1_1349@runat_14022003_100733_2299 -NCPV=6 -ICPV=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description |
|------------|-------|---------|--------|-------|-------------|
| 1 | 84.5 | 14.7 | 504 | 10 | BE581359 |
| 2 | 83 | 14.5 | 704 | 17 | AQ643182 |
| 3 | 82 | 14.3 | 460 | 12 | BP141291 |
| 4 | 82 | 14.3 | 470 | 12 | BP137845 |
| 5 | 82 | 14.3 | 484 | 12 | BP141201 |
| 6 | 82 | 14.3 | 493 | 12 | BP141314 |
| 7 | 82 | 14.3 | 496 | 12 | BP139798 |
| 8 | 82 | 14.3 | 529 | 12 | BP136045 |
| 9 | 82 | 14.3 | 541 | 10 | AW737851 |
| 10 | 82 | 14.3 | 548 | 12 | BP139905 |
| 11 | 82 | 14.3 | 561 | 12 | BP136094 |
| 12 | 82 | 14.3 | 561 | 12 | BP138707 |
| 13 | 82 | 14.3 | 561 | 12 | BP139936 |
| 14 | 82 | 14.3 | 561 | 12 | BP140418 |
| 15 | 82 | 14.3 | 561 | 12 | BP140421 |
| 16 | 82 | 14.3 | 567 | 12 | BP139637 |
| 17 | 82 | 14.3 | 597 | 12 | BP136116 |
| 18 | 82 | 14.3 | 600 | 10 | AW738722 |
| 19 | 82 | 14.3 | 620 | 10 | AW218016 |
| 20 | 82 | 14.3 | 648 | 12 | BP140407 |
| 21 | 82 | 14.3 | 652 | 12 | BP136420 |
| 22 | 82 | 14.3 | 663 | 10 | AW738268 |
| 23 | 82 | 14.3 | 672 | 12 | BP137992 |
| 24 | 82 | 14.3 | 687 | 12 | BP136406 |
| 25 | 82 | 14.3 | 697 | 10 | AW738284 |
| 26 | 82 | 14.3 | 693 | 13 | BI933236 |
| 27 | 82 | 14.3 | 706 | 13 | BI933320 |
| 28 | 82 | 14.3 | 723 | 13 | BI933312 |
| 29 | 80.5 | 14.0 | 499 | 12 | BP138008 |
| 30 | 80 | 14.0 | 459 | 12 | BP138662 |
| 31 | 79.5 | 13.9 | 656 | 12 | BF708357 |
| 32 | 79.5 | 13.9 | 1058 | 17 | CNS05WUS |
| 33 | 79 | 13.8 | 470 | 12 | BP138190 |
| 34 | 78 | 13.6 | 566 | 13 | BJ101346 |
| 35 | 78 | 13.6 | 947 | 17 | AZ676398 |
| 36 | 77 | 13.4 | 410 | 17 | AA549977 |
| 37 | 77 | 13.4 | 663 | 17 | BP136991 |
| 38 | 76 | 13.3 | 508 | 17 | AQ651585 |
| 39 | 76 | 13.3 | 721 | 9 | AL726177 |
| 40 | 76 | 13.3 | 1051 | 10 | BE259556 |
| 41 | 75.5 | 13.2 | 473 | 12 | BG086881 |
| 42 | 75.5 | 13.2 | 478 | 13 | BI174761 |
| 43 | 75.5 | 13.2 | 526 | 17 | AO701970 |
| 44 | 75.5 | 13.2 | 640 | 13 | BJ122797 |
| 45 | 75.5 | 13.2 | 643 | 10 | BB645658 |

ALIGNMENTS

RESULT 1
LOCUS BE581359
DEFINITION BE581359 504 bp mRNA linear EST 09-MAY-2001
LOCUS BE581359
SEQUENCE BE581359
ACCESSION BE581359.1 GI:9832301
VERSION
KEYWORDS
SOURCE
ORGANISM Strongyloides stercoralis.
Strongyloides stercoralis.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae;
Panagrolaimoidae; Strongyloidea; Strongyloidea.
REFERENCE
1 (bases 1 to 504)
McCartier, V., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.

Gibbons, M., Rutter, E., Bennett, J., Franklin, C., Teagareishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harrey, N., Schurk, R., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCan, K., Waterston, R., and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
High quality sequence stop: 414.
Location/Qualifiers
1..504

FEATURES

source

/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TBN95TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)
/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site.1: EcoRI; Site.2: XhoI; mRNA was purified from 2 x 10E3 rhabditiform larvae which had been isolated from gerbils experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range: 100-1700)."
BASE COUNT 200 a 62 c 59 g 183 t
ORIGIN

Alignment Scores:

Pred. No.: 0.754 Length: 504
Score: 84.50 Matches: 19
Percent Similarity: 54.908 Conservative: 9
Best Local Similarity: 37.258 Mismatches: 19
Query Match: 14.75% Indels: 4
DB: 10 Gaps: 2

US-09-936-737a-2 (1-103) x BE581359 (1-504)

QY 9 PheTYrAlAsnArGlySTYrThr-AspPheAspLysSerPheLysLysSerSerAspLeu 28
Db 319 TGGTACACAAATAGGAAAGACAGATATATCACTAGTAATGAGAACTAGATT 378
QY 28 uAspGlyLysLysLysThrcysPheLysThrGluTYrCysTYr---lLeValPheGluAs 47
Db 379 AAAAGAAATGATTAACAAAGTATGCAATTTTGTATTCATTAATTAATGATGA 438
QY 47 PthValAsnLysGlyLys-----TYrTYr 55
Db 439 AACATTAATAAATCGTTGTCTTTTATTAT 469
RESULT 2
LOCUS A0643182 704 bp DNA linear GSS 08-JUL-1999
DEFINITION RPI93-ECORI-6L8.TJ RPI93-ECORI Trypanosoma brucei genomic clone
ACCESSION A0643182
VERSION A0643182.1 GI:5119892
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE

1 (bases 1 to 704)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C., and Adams, M.

TITLE

JOURNAL

COMMENT

Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPI93 Unpublished (1999)
Library for gene discovery and sequence-ready map construction
Other GSSs: RPI93-ECORI-6L8.TJ
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search page: <http://www.tigr.org/ldb/ldb/tbdb/>.
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..704

FEATURES

source

/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="RPI93-ECORI-6L8"
/note="Vector: pBACe3.6; Site.1: Eco RI; Site.2: Eco RI; constructed for The Institute for Genomic Research by Bohui Zhao in Pieter de Jong's laboratory (Roswell Park Cancer Institute, Buffalo, NY). Briefly, Trypanosoma brucei TREU927/4 GUTat 10.1 agarose embedded DNA was partially digested with a combination of Eco RI and Eco RI methylase (RPI93-ECORI segment) or Dpn II (RPI93-DpnII segment). High molecular weight fragments were ligated in pBACe3.6 vector digested with Eco RI or Bam HI, respectively. The average insert size is 141 Kb. Total coverage (both segments): > 90 x the haploid non-mitochondrial genome."
BASE COUNT 290 a 92 c 147 g 175 t
ORIGIN

Alignment Scores:

Pred. No.: 1.77 Length: 704
Score: 83.00 Matches: 26
Percent Similarity: 40.43% Conservative: 12
Best Local Similarity: 27.66% Mismatches: 24
Query Match: 14.49% Indels: 32
DB: 17 Gaps: 3

US-09-936-737a-2 (1-103) x A0643182 (1-704)

QY 11 AlaAsnAlaGlySTYrThrAspPheAspLysSerPheLysLysSerSerAspLeuAspLeu 30
Db 70 GCAACCCCGAGACAGAACAGAACTTCGAAAGAAAGAAAGAGAGATATA 129
QY 31 CysLysLysThrcysPheLysThrcysTYrLleValPheGluAspPthValAsn 50
Db 130 TGCATA----- 135
QY 51 LysGlyLysTYrTYrAsnValAlaAspGlyGluLeuAspGlnGluLysPheValVal 70
Db 136 -----GATGGC-----TCCAAGATGGGTTCAATGG 162
QY 71 AspGluAsnPheThrcysAsnTYrLleuThraspCysGlyGlyLysAsp-----Ala 87
Db 163 GATGAGAAATGCACATGACAAAGCAAGGAGCTGCAAACTAAAGAAAGAAAGAAAGACA 222
QY 88 GlyAsnAlaAlaGlyThrcysAspGluSerAspGluValAsp 101
Db 223 AACACAGCAGCAGAGACAGAGACAGATACAGGGGTATGAT 264
RESULT 3

| | | | | | |
|---|--|--------------------------------|------|--------|-----------------|
| LOCUS | BG141291 | 462 bp | mRNA | linear | EST 31-JAN-2001 |
| DEFINITION | EST481733 wild tomato pollen Lycopersicon pennellii cDNA clone | | | | |
| ACCESSION | CLPP20L16 5' sequence, mRNA sequence. | | | | |
| VERSION | BG141291 | | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | Lycopersicon pennellii. | | | | |
| ORGANISM | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon. | | | | |
| REFERENCE | 1 (bases 1 to 462) van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T., Hansen,C., Rönning,C. and Tanksley,S. Generation of ESTs from wild tomato (L. pennellii) pollen Unpublished (2001) | | | | |
| TITLE | Contact: CUGI | | | | |
| JOURNAL | Clemson University Genomics Institute | | | | |
| COMMENT | Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..462 /organism="Lycopersicon pennellii" /cultivar="TA56" /db_xref="taxon:28526" /clone="CLPP20L16" /clone_1lb="wild tomato pollen" /tissue_type="pollen" /dev_stage="pollen collected from open flowers" /lab_host="SOLR" /note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2: XhoI; Pollen was collected from open flowers from L.pennellii TA56, and stored at -80 C until library construction." | | | | |
| BASE COUNT | 175 a | 85 c | 68 g | 134 t | |
| ORIGIN | | | | | |
| Alignment Scores: | | | | | |
| Pred. No.: | 1..42 | Length: | 462 | | |
| Score: | 82.00 | Matches: | 26 | | |
| Best Similarity: | 48.68% | Conservative: | 11 | | |
| Best Local Similarity: | 34.21% | Mismatches: | 27 | | |
| Query Match: | 14.31% | Indels: | 12 | | |
| DB: | 12 | Gaps: | 4 | | |
| US-09-936-737A-2 (1-103) x BG141291 (1-462) | | | | | |
| QY 14 | LysTyrThrAspPhe----- | AspLysSerPheLysLysSerSerAspLeu | 28 | | |
| DB 238 | AAAGACACAAAGATTTTATCAAAAATGTAATTCATATAAAGATTGGCTTCAGTCACCAATA | 111111 111111 | 297 | | |
| QY 29 | AspGluCysLysLysLysThrCysPheLysThrGluTyrCysTyrLleValPheGluAspThr | 48 | | | |
| DB 298 | GATCATATATAGAAAGAAAGATTGTTTAAAAACA-----TGCAAAGAGAGGTATGAAAGATGCA | 351 | | | |
| QY 49 | ValAsn-----LysGluCysTyrTyrAsnValValAspGluGluLeuAspGluGlu | 66 | | | |
| DB 352 | ATTGATTCATATGAAAAAGGCAACACAAAGATT-----AAAGACTTAATTAATAC | 402 | | | |
| QY 67 | LysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCys | 82 | | | |
| DB 403 | AAAGCAATATGATGATATTGTCATTCATATACATACCTTGACACTTGT | 450 | | | |
| RESULT 4 | | | | | |
| LOCUS | BG137845 | 470 bp | mRNA | linear | EST 31-JAN-2001 |
| DEFINITION | EST478287 wild tomato pollen Lycopersicon pennellii cDNA clone | | | | |
| ACCESSION | CG137845 5' sequence, mRNA sequence. | | | | |
| VERSION | BG137845 | 1 GI:12638033 | | | |

| KEYWORDS | EST. | | |
|---|---|---------------|------|
| SOURCE | Lycopersicon pennellii. | | |
| ORGANISM | Lycopersicon pennellii | | |
| | Eukaryote: Vitridipnatae; Streptophyta; Embryophyta; Tracheophyta; | | |
| | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; | | |
| | Asteridae: euasterids I; Solanales; Solanaceae; Solanum; | | |
| | Lycopersicon. | | |
| REFERENCE | 1 (bases 1 to 470) | | |
| AUTHORS | van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T., Hansen, C., Ronting, C. and Tanksley, S. | | |
| TITLE | Generation of ESTs from wild tomato (L. pennellii) pollen | | |
| JOURNAL | Unpublished (2001) | | |
| COMMENT | Contact: CUGI | | |
| | Clemson University Genomics Institute | | |
| | Clemson University | | |
| | 100 Jordan Hall, Clemson, SC 29634, USA | | |
| | Email: http://www.genome.clemson.edu/orders/index.html . | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..470 | | |
| | /organism="Lycopersicon pennellii" | | |
| | /cultivar="TA56" | | |
| | /db_xref="taxon:28526" | | |
| | /clone="CLP7D15" | | |
| | /clone_lib="wild tomato pollen" | | |
| | /issue_type="pollen" | | |
| | /dev_stage="pollen collected from open flowers" | | |
| | /lab_host="SOLR" | | |
| | /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L.pennellii TA56, and stored at -80 C until library construction." | | |
| BASE COUNT | 177 a 86 c 70 g 137 t | | |
| ORIGIN | | | |
| Alignment Scores: | | | |
| Pred. No.: | 1.45 | Length: | 470 |
| Score: | 82.00 | Matches: | 26 |
| Percent Similarity: | 48.68% | Conservative: | 11 |
| Best Local Similarity: | 34.21% | Mismatches: | 27 |
| Query Match: | 14.31% | Indels: | 12 |
| DB: | 12 | Gaps: | 4 |
| US-09-936-737A-2 (1-103) x BG137845 (1-470) | | | |
| QY | 14 LysTyrThrAsphe-----AspLysSerPheLysSerSeraspLeu 28 | | |
| | | | |
| DB | 238 AAGACACAGAGATTATTCAAAAATGTAATGATATAAAGTTGGCTTCAGTCACCAATA 297 | | |
| QY | 29 AspGluCysLysLysThrCysPheLysThrGluTyrCysTyrTleValPheGluaspThr 48 | | |
| | | | |
| DB | 298 GATCATATATATGAAGAAGATGTTTAAACA-----TGCAAAGAGGTGTATGAAGATGCA 351 | | |
| QY | 49 ValAsn-----LysGluCysTyrTyrAsnValValAspGluGluLeuaspGlnGlu 66 | | |
| | | | |
| DB | 352 ATTGATTCAATGAAAAAGCACACAGAGTT-----AAAGCAATAATATATTAC 402 | | |
| QY | 67 LysPheValValaspGluasnPheThrGluAsnTyrLeuThrAspCys 82 | | |
| | | | |
| DB | 403 AAAGCAATATGCTATTGTCATCATATATACATTGACACTTGT 450 | | |
| RESULT 5 | | | |
| LOCUS | BG141201 | 484 bp | mRNA |
| DEFINITION | EST461643 wild tomato pollen Lycopersicon pennellii cDNA clone | | |
| | CLP20110 5' sequence, mRNA sequence. | | |
| ACCESSION | BG141201 | | |
| VERSION | BG141201.1 | | |
| KEYWORDS | EST. | | |
| SOURCE | Lycopersicon pennellii. | | |
| ORGANISM | Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | |
| | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; | | |
| | Asteridae: euasterids I; Solanales; Solanaceae; Solanum; | | |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|--|--|---|--|
| 1 (bases 1 to 496) | van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T., Hansen, C., Roming, C., and Tanksley, S. | Generation of ESTs from wild tomato (<i>L. pennellii</i>) pollen | Unpublished (2001) | Contact: CUGI |
| | Clemson University Genomics Institute | Clemson University | 100 Jordan Hall, Clemson, SC 29634, USA | Email: http://www.genome.clemson.edu/orders/index.html . |
| | | | Location/Qualifiers | |

Xhol: supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA96). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT

196 a 96 c 90 g 159 t

Alignment Scores:

| Pred. No.: | 1.72 | Length: | 541 |
|------------------------|--------|---------------|-----|
| Score: | 82.00 | Matches: | 26 |
| Percent Similarity: | 48.68% | Conservative: | 11 |
| Best Local Similarity: | 34.21% | Mismatches: | 27 |
| Query Match: | 14.31% | Indels: | 12 |
| DB: | 10 | Gaps: | 4 |

US-09-936-737A-2 (1-103) x AM737851 (1-541)

OY 14 LysTyrThrAspPhe-----AspLysSerPheLysSerSerAspLeu 28
 Db 225 AAGACACAGATTTTATCAAGAAATGTAATGATATAAGATTGGCTTCAAGTCACCAATA 284
 OY 29 AspLysCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr 48
 Db 285 GATCATTAACAAAAAGATTGTTAAACA-----TGCAGAGAGGTGTATGAGATGCA 338
 OY 49 ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu 66
 Db 339 ATTGATTCATGAAAGAACACACAGATGTT-----AAGAACTAAATATATATAC 389
 OY 67 LysPheValValAspLysAsnPheThrGluAsnTyrLeuThrAspCys 82
 Db 390 AAGCAAAATATGATATGTTGTCATCATATACATACCTTGACACTTGT 437

RESULT 10

LOCUS

BG139905 548 bp mRNA linear EST 31-JAN-2001

DEFINITION

EST480347 wild tomato pollen Lycopersicon pennellii cDNA clone

ACCESSION

BG139905

VERSION

BG139905.1 GI:12640093

KEYWORDS

EST

SOURCE

ORGANISM

Lycopersicon pennellii

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 548)

van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T., Hansen, C., Roming, C. and Tanksley, S.

Generation of ESTs from wild tomato (L. pennellii) pollen

Unpublished (2001)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

Location/Qualifiers

1.548

/organism="Lycopersicon pennellii"

/cultivar="TA56"

/db_xref="taxon:28526"

/clone="CLP14N18"

/clone_lib="wild tomato pollen"

/tissue_type="pollen"

/dev_stage="pollen collected from open flowers"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction."

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

199 a 99 c 87 g 163 t

Alignment Scores:

| Pred. No.: | 1.75 | Length: | 548 |
|------------------------|--------|---------------|-----|
| Score: | 82.00 | Matches: | 26 |
| Percent Similarity: | 48.68% | Conservative: | 11 |
| Best Local Similarity: | 34.21% | Mismatches: | 27 |
| Query Match: | 14.31% | Indels: | 12 |
| DB: | 12 | Gaps: | 4 |

US-09-936-737A-2 (1-103) x BG139905 (1-548)

OY 14 LysTyrThrAspPhe-----AspLysSerPheLysSerSerAspLeu 28
 Db 225 AAGACACAGATTTTATCAAAATGTAATGATATAAGATTGGCTTCAAGTCACCAATA 284
 OY 29 AspLysCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr 48
 Db 285 GATCATTAACAAAAAGATTGTTAAACA-----TGCAGAGAGGTGTATGAGATGCA 338
 OY 49 ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu 66
 Db 339 ATTGATTCATGAAAGAACACACAGATGTT-----AAGAACTAAATATATATAC 389
 OY 67 LysPheValValAspLysAsnPheThrGluAsnTyrLeuThrAspCys 82
 Db 390 AAGCAAAATATGATATGTTGTCATCATATACATACCTTGACACTTGT 437

RESULT 11

LOCUS

BG136094 561 bp mRNA linear EST 31-JAN-2001

DEFINITION

EST476536 wild tomato pollen Lycopersicon pennellii cDNA clone

ACCESSION

BG136094

VERSION

BG136094.1 GI:12636282

KEYWORDS

EST

SOURCE

ORGANISM

Lycopersicon pennellii

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 561)

van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T., Hansen, C., Roming, C. and Tanksley, S.

Generation of ESTs from wild tomato (L. pennellii) pollen

Unpublished (2001)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

Location/Qualifiers

1.561

/organism="Lycopersicon pennellii"

/cultivar="TA56"

/db_xref="taxon:28526"

/clone="CLP1J21"

/clone_lib="wild tomato pollen"

/tissue_type="pollen"

/dev_stage="pollen collected from open flowers"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction."

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

208 a 99 c 90 g 164 t

BASE COUNT

199 a 99 c 87 g 163 t

ORIGIN

1.8

Length:

561

Matches:

26

Conservative:

11

Mismatches:

27

Indels:

12

Gaps:

4

Alignment Scores:

Pred. No.:

1.8

Length:

561

Matches:

26

Conservative:

11

Mismatches:

27

Indels:

12

Gaps:

4

Alignment Scores:

Pred. No.:

1.8

Length:

561

Matches:

26

Conservative:

11

Mismatches:

27

Indels:

12

| | | | |
|--------------|--------|---------|----|
| Query Match: | 14.31% | Indels: | 12 |
| DB: | 12 | Gaps: | 4 |

US-09-936-737A-2 (1-103) x BG136094 (1-561)

| | | | |
|----|-----|---|-----|
| Oy | 14 | LysTyrThrAspPhe-----AspYSerPheLysSerSerAspLeu | 28 |
| | | | |
| Dd | 238 | AAGACACAAGATTTTATCAAAAATGCTATTGATAAAAGATTGGCTTCAGAGCACCATTA | 297 |
| Oy | 29 | AspGluCysLysTsrHrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr | 48 |
| | | | |
| Dd | 298 | GATCATTTAAGAAAAAGATTGTTTTAAACA-----TCCAAAGAAGTGATGAACATCA | 351 |
| Oy | 49 | ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu | 66 |
| | | | |
| Dd | 352 | ATTGATTCAATGAAAAAGAACACACAGATGTT-----AAAGACTTAATATTATAC | 402 |
| Oy | 67 | LysPheValValAspGluAsnPheThrGluAsnTyrIleThrAspCys | 82 |
| | | | |
| Dd | 403 | AAAGCAATATGCGATTATTTGGTCATCATCATATACATACCCTTGACACATTTGT | 450 |

RESULT 12
BG138707 561 bp mRNA linear EST 31-JAN-2001
LOCUS BG138707
DEFINITION EST479149 wild tomato pollen Lycopersicon pennellii cDNA clone
cclp10f12 5' sequence, mRNA sequence.
ACCESSION BG138707 GI:12638895
VERSION BG138707.1 GI:12638895
KEYWORDS EST.
SOURCE Lycopersicon pennellii.
ORGANISM Lycopersicon pennellii.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 561)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Rønning,C. and Tanksley,S.
Generation of ESTs from wild tomato (L. pennellii) pollen
Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES
source location/Qualifiers
1..561
/organism="Lycopersicon pennellii"
/cultivar="FA56"
/db_xref="taxon:28526"
/clone="CLP10F12"
/clone_lib="wild tomato pollen"
/tissue_type="pollen"
/dev_stage="pollen collected from open flowers"
/lab_host="SOLR"
/note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
XhoI. Pollen was collected from open flowers from
L.pennellii FA56, and stored at -80 C until library
construction."

| | | | |
|------------|-------|------|-------|
| BASE COUNT | 208 a | 99 c | 164 t |
| ORIGIN | | | |

Alignment Scores:

| | | | |
|------------------------|--------|---------------|-----|
| Pred. NO.: | 1 8 | Length: | 561 |
| Percent Similarity: | 82.00 | Matches: | 26 |
| Best Local Similarity: | 48.08% | Conservative: | 11 |
| Query Match: | 34.21% | Mismatches: | 27 |
| | 14.31% | Indels: | 12 |
| DB: | 12 | Gaps: | 4 |

US-09-936-737A-2 (1-103) x BG138707 (1-561)

14 LysTyrThrAspPhe-----AspYSerPheLysSerSerAspLeu 28

| DB | LOCUS | DEFINITION | VERSION | KEYWORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE | COMMENT | FEATURES |
|--|--|--|---------|----------|-----------------|----------|-----------|---------|-------|---------|----------|
| Db | 238 | AAAGACACAAGATTATTATCATCAAAAATGTAAATTCATTAAGAGATTGGCTTCAAGTGACCAATA | 297 | | | | | | | | |
| Qy | 29 | AspGluCysLysValSerThrCysPheLysThrGluTrpCysTyrIleValPheGluAspThr | 48 | | | | | | | | |
| Db | 298 | GATCATTTATTAAGAAAGATTGTTTAAACA-----TGCAAAGAGGCGTATGAAGTGA | 351 | | | | | | | | |
| Qy | 49 | ValAsn-----LysGluCysTyrTyrAsnValValAspGluGluLeuAspGlnGlu | 66 | | | | | | | | |
| Db | 352 | ATTGATTCAATGAAGAAAAAGCAACACAGATGTT-----AAAGACTAAATTTATAC | 402 | | | | | | | | |
| Qy | 67 | LysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCys | 82 | | | | | | | | |
| Db | 403 | AAAGCAATATATGATTTGTCGTCATATACATACACTTGACACTTGT | 450 | | | | | | | | |
| RESULT 13 | | | | | | | | | | | |
| LOCUS | BGI139936 | 561 bp | MRNA | linear | EST 31-JAN-2001 | | | | | | |
| DEFINITION | EST480378 wild tomato pollen Lycopersicon pennellii cDNA clone | | | | | | | | | | |
| ACCESSION | CLIP153 5' sequence, mRNA sequence. | | | | | | | | | | |
| VERSION | BGI139936.1 GI:12640124 | | | | | | | | | | |
| KEYWORDS | EST. | | | | | | | | | | |
| SOURCE | Lycopersicon pennellii. | | | | | | | | | | |
| ORGANISM | Lycopersicon pennellii; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon. | | | | | | | | | | |
| REFERENCE | 1 (bases 1 to 561) | | | | | | | | | | |
| AUTHORS | van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T., Hansen, C., Roming, C., and Tanksley, S. | | | | | | | | | | |
| TITLE | Generation of ESTs from wild tomato (L. pennellii) pollen | | | | | | | | | | |
| COMMENT | Unpublished (2001) | | | | | | | | | | |
| | Contact: CUGI | | | | | | | | | | |
| | Clemson University Genomics Institute | | | | | | | | | | |
| | Clemson University | | | | | | | | | | |
| | 100 Jordan Hall, Clemson, SC 29634, USA | | | | | | | | | | |
| | Email: http://www.genome.clemson.edu/orders/index.html . | | | | | | | | | | |
| | Location/Qualifiers | | | | | | | | | | |
| FEATURES | 1..561 | | | | | | | | | | |
| source | /organism="Lycopersicon pennellii" | | | | | | | | | | |
| | /cultivar="TA56" | | | | | | | | | | |
| | /db_xref="taxon:28526" | | | | | | | | | | |
| | /clone="CLIP153" | | | | | | | | | | |
| | /clone_lib="wild tomato pollen" | | | | | | | | | | |
| | /tissue_type="pollen" | | | | | | | | | | |
| | /dev_stage="pollen collected from open flowers" | | | | | | | | | | |
| | /lab_host="SOLR" | | | | | | | | | | |
| | /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction." | | | | | | | | | | |
| BASE COUNT | 208 a 99 c 165 t | | | | | | | | | | |
| ORIGIN | | | | | | | | | | | |
| Alignment Scores: | | | | | | | | | | | |
| Pred. No.: | 1.8 | length: | 561 | | | | | | | | |
| Score: | 82.00 | Matches: | 26 | | | | | | | | |
| Percent Similarity: | 48.68% | Conservative: | 11 | | | | | | | | |
| Best Local Similarity: | 34.21% | Mismatches: | 27 | | | | | | | | |
| Query Match: | 14.31% | Indels: | 12 | | | | | | | | |
| DB: | 12 | Gaps: | 4 | | | | | | | | |
| US-09-936-737A-2 (1-103) x BGI139936 (1-561) | | | | | | | | | | | |
| Qy | 14 | LysTyrThrAspPhe-----AspLysSerPheLysSerSerAspLeu | 2 | | | | | | | | |

